Réponse à l'infection : apport du transcriptome

Julien Textoris

UNIVERSITÉ DE LA MÉDITERRANÉE

FACULTE DE MÉDECINE DE MARSEILLE

ECOLE DOCTORALE : Sciences de la vie et de la santé

TITRE DE LA THÈSE :

Réponse à l'infection : apport du transcriptome

T H È S E

Présentée et publiquement soutenue devant

LA FACULTÉ DE MÉDECINE DE MARSEILLE

Le 30 juin 2011

Par M Textoris Julien
né le 10 mars 1979 à Ingwiller

Pour obtenir le grade de DOCTEUR de L'UNIVERSITÉ de la MÉDITERRANÉE

SPÉCIALITÉ : Pathologie humaine, maladies infectieuses

Membres du Jury de la Thèse :

Professeur Jean-Louis Mège ….....	Président
Professeur Didier Payen de La Garanderie.....	Rapporteur
Professeur Jean-Paul Mira ….....	Rapporteur
Professeur Marc Leone ….....	Examinateur
Docteur Catherine Nguyen ….....	Examinateur
Docteur Jean-Jacques Diaz ….....	Examinateur

Table des matières

Avant propos

L'objectif de cette thèse est d'explorer l'inflammation et l'infection au niveau du transcriptome, à l'aide de la technologie des puces à ADN.

Pour cela, nous avons dans un premier temps travaillé sur des données publiques. Nous avons construit une base de données de signatures transcriptionnelles annotées, et développé un logiciel modulaire d'analyse. Ce logiciel permet d'explorer aisément les données publiques en effectuant des recherches par nom de gène ou par mots-clés. Nous avons ensuite exploré la modulation temporelle de l'expression des gènes du parenchyme pulmonaire dans un modèle murin d'inflammation aiguë par injection d'acide oléique. Dans un second modèle murin d'infection par *Coxiella burnetii*, nous avons analysé le rôle du sexe dans la modulation de la réponse transcriptionnelle hépatique, et identifié des voies métaboliques impliquées dans le contrôle de l'infection. Dans un troisième modèle *in-vitro* d'infection par différentes souches du virus de la grippe, nous avons identifié une signature transcriptionnelle commune de réponse à l'infection. Par une approche bio-informatique originale, cette signature a conduit à l'identification de nouveaux anti-viraux à large spectre, dont l'efficacité a été démontrée *in-vitro* sur les souches utilisées pour l'analyse, et sur la souche H1N1, responsable de la dernière pandémie grippale. Enfin, nous avons analysé les modulations du transcriptome lors de pneumonies associées à la ventilation mécanique compliquant l'évolution de sujets traumatisés graves admis en réanimation.

Introduction

Le sepsis est un syndrome clinique qui décrit une infection grave associée à un syndrome inflammatoire généralisé (« SIRS » pour « Systemic Inflammatory Response Syndrome » en anglais). Il est caractérisé par une agression aiguë de l'hôte par des médiateurs dont l'expression est induite par l'infection. Son incidence augmente avec le temps du fait des progrès de la médecine (vieillissement de la population), mais la mortalité reste aux alentours de 35 % [1]. D'un point de vue clinique, le patient a généralement des signes aspécifiques comme une fièvre ou une hypothermie, une tachypnée, une tachycardie et une hyperleucocytose ou une leucopénie. Le SIRS ainsi défini recouvre des situations pathologiques variées, souvent sans relation avec une infection.

Le sepsis se complique souvent d'une ou plusieurs défaillances d'organe incluant une hypotension, une insuffisance cardiaque, un coma, une insuffisance rénale ou une coagulation intravasculaire disséminée. Il est alors désigné sous le terme de sepsis grave. Lorsque la défaillance cardio-vasculaire prédomine et qu'il existe une hypotension qui ne répond pas au remplissage vasculaire, le terme de choc septique est employé. Il existe un continuum nosologique entre le SIRS, le sepsis, le sepsis grave et le choc septique (Figure 1). Les formes graves sont associées à une mortalité d’environ 50 % dans le choc septique [2].

D'un point de vue physiopathologique, le sepsis est un phénomène complexe, où l'infection initiale active une réponse systémique de l'organisme, responsable des défaillances d'organe et de la mortalité. C'est un phénomène dynamique où des

molécules du micro-organisme (endotoxine par exemple) entraînent la synthèse de diverses cytokines, chemokines, radicaux libres, etc … La réponse de l'organisme met en jeu l'immunité innée et adaptative (figure 2), en étroite relation avec les réponses de l'endothélium vasculaire et du système hormonal [3,4].

Pour comprendre cette physiopathologie complexe, un modèle constitué de deux phases de réponse a été proposé. Le sepsis consisterait en une phase pro-inflammatoire initiale suivie d'une phase de réponse compensatrice anti-inflammatoire [5] (figure 3). La recherche pharmaceutique s'est longtemps focalisée sur la seule réponse pro-inflammatoire excessive. Malheureusement, en dépit de résultats positifs dans des modèles animaux [6-8], les essais cliniques n'ont pas démontré l'efficacité de cette approche. Ce modèle a donc été progressivement remis en cause.

Pour comprendre la complexité de ce syndrome au lit du patient, des cliniciens ont cherché à déterminer des biomarqueurs spécifiques du sepsis. Les biomarqueurs évalués sont le plus souvent des molécules de la réponse pro-inflammatoire. Le dosage de diverses cytokines chez les patients septiques, a permis de mieux appréhender la physiopathologie du sepsis. Toutefois, ceci ne s'est pas concrétisé par des progrès significatifs en termes de prise en charge des patients. De plus, de nombreuses études soulignent le caractère contradictoire de ces résultats [9-11]. Enfin, malgré des travaux relativement homogènes dans des modèles animaux stéréotypés, il est souvent difficile de déterminer en clinique le statut pro- ou anti-inflammatoire d'un patient sur le seul dosage de quelques cytokines. En effet, le résultat dépend souvent du choix des cytokines et du moment du prélèvement.

En réanimation, une des complexités de prise en charge des patients est liée à l'incapacité de prédire la nature infectieuse ou non d'un état inflammatoire. Pour évaluer l'état des patients, les réanimateurs disposent de nombreux outils intégrant des variables cliniques et biologiques à l'admission, comme l'indice de gravité simplifié (IGS) II [12] ou le score APACHE II [13]. Toutefois, si ces outils sont adaptés pour traiter des données collectives, leur performance est moindre à l'échelon individuel. Depuis une vingtaine d'années, de nombreuses études ont été conduites afin de pouvoir : *(1)* mesurer une susceptibilité individuelle d'origine héréditaire ; *(2)* identifier un profil complexe de réponse biologique permettant d'affiner le diagnostic ; *(3)* utiliser la modulation de cette réponse biologique pour suivre la réponse au traitement. Les hypothèses biologiques qui soutiennent ces approches sont que : *(1)* la susceptibilité à une maladie est en partie déterminée par notre patrimoine génétique ; *(2)* la réponse de l'hôte à l'agression et aux traitements mis en œuvre peut être prédite au niveau moléculaire. Ainsi, le clinicien espère utiliser l'outil moléculaire pour affiner son diagnostic, classer les patients en fonction de leur gravité et de leur pronostic, et suivre la réponse aux traitements mis en œuvre.

Les progrès technologiques permettent d'envisager l'analyse globale de l'expression du génome au niveau du transcriptome [14]. En effet, cette analyse globale permet de s'affranchir du biais de sélection *a priori* des études de dosage de cytokines, où seuls certains gènes pré-sélectionnés sont explorés.

Qu'est-ce que le transcriptome ?

Le transcriptome est un terme générique représentant un ensemble de molécules d'acide ribonucléique (ARN). Ces ARN sont transcrits à partir du génome. Les transcriptomes diffèrent selon le type cellulaire, le moment du recueil et les stimulations reçues par la cellule. Le génome, qui est composé d'acide désoxyribonucléique (ADN), est identique dans l'ensemble des cellules d'un organisme. Par contre, son expression varie au cours du temps. Par conséquent, le transcriptome est une photographie « instantanée » de l'expression du génome. Par extension, le terme transcriptome est utilisé pour les ARN extraits à partir d'un type cellulaire ou d'un tissu.

Le transcriptome regroupe l'ensemble des molécules d'ARN transcrites à partir du génome [15]. Cela comprend les ARN messagers (ARNm), qui sont traduits par les ribosomes en protéines, et les ARN ribosomiques, les ARN de transfert, et tous les autres ARN non codant [16] qui assurent diverses fonctions au sein de la cellule. L'expression du génome met en jeu un grand nombre de mécanismes de modifications post-transcriptionnelles des ARN. Par conséquent, l'ensemble théorique des ARN synthétisés à partir du génome est plus vaste que le génome lui-même. Une des caractéristiques du transcriptome est donc de varier en fonction des conditions et au cours du temps.

Méthodes d'analyse du transcriptome

Il existe de nombreuses méthodes d'analyse du transcriptome : analyse basée sur le type d'ARN étudiés, analyse d’un ou de quelques ARN messagers (par PCR,

« Polymerase Chain Reaction » en anglais), ou analyse d'un ensemble d'ARN.

Historiquement, la comparaison de deux transcriptomes faisait appel à la construction d'une banque soustractive (figure 4) [17]. Cette méthode est basée sur la propriété d'hybridation complémentaire des acides nucléiques. Elle est qualitative et non exhaustive. Elle identifie un certain nombre de gènes dont l'expression varie entre les deux transcriptomes étudiés (par exemple T_A et T_B, figure 4). Pour cela, les ARNm sont rétro-transcrits en ADN complémentaires (ADNc) à double-brin. L'un des deux transcriptomes est rétro-transcrit en incorporant des nucléotides marqués (T_B, par exemple par de la biotine). Les ADNc des deux transcriptomes sont alors mélangés afin de permettre l'hybridation des ADNc identiques dans les deux transcriptomes. Les ADNc du transcriptome marqué (T_B), ainsi que les ADNc hybrides entre les deux transcriptomes sont alors éliminés en utilisant un ligand (streptavidine si les nucléotides sont marqués à la biotine). Il ne reste donc que les ADNc des gènes exprimés uniquement dans le transcriptome A. Ces ADNc sont clonés dans des vecteurs d'expression, puis isolés après transformation dans une bactérie. Les clones d'intérêts sont alors cultivés. Le gène d'intérêt est purifié et séquencé pour identification.

Les techniques plus récentes d'analyse du transcriptome permettent une analyse quantitative de l'expression des gènes. La technique SAGE (pour « *Serial Analysis of Gene Expression* ») est basée sur le séquençage (figure 5) [18]. Les ARNm sont rétro-transcrits en ADNc, puis ces derniers sont soumis à une digestion par un enzyme de restriction. Les fragments obtenus sont ligaturés les uns aux autres et clonés dans un

vecteur de séquençage. On obtient des fragments d'ADNc juxtaposés et séparés par un motif connu (le *site* de l'enzyme de restriction). Le long fragment hybride est séquencé, puis, par analyse bioinformatique, les séquences de chaque fragment sont extraites. Les fragments identiques sont comptabilisés. L'analyse nécessite une étape d'annotation pour identifier le gène dont les fragments sont issus, par alignement sur un génome de référence. Cette technique permet d'obtenir pour chacun des gènes identifiés un compte du nombre de fragments clonés et séquencés.

Les puces à ADN (*DNA microarrays*) sont un outil d'analyse quantitative de l'expression des gènes (figure 6) permettant l'analyse de l'expression de plusieurs milliers de gènes en une seule expérience [20]. La puce à ADN est généralement une membrane de nylon ou une lame de verre sur laquelle sont attachés des morceaux d'ADN (sondes) dont la séquence est spécifique d'un gène donné. Les sondes sont soit synthétisées *in-situ* soit déposées par un robot selon un motif régulier et connu. Les sondes sont soit des oligonucléotides (d'une longueur de quelques dizaines de nucléotides), soit des ADNc (d'une longueur de quelques centaines de nucléotides). Les ARNm de l'échantillon étudié sont marqués à l'aide de nucléotides associés à une molécule radioactive ou un fluorochrome au cours d'une étape de rétro-transcription. La sensibilité de l'outil évite normalement une étape d'amplification des ADNc. Selon le même principe d'hybridation spécifique des séquences complémentaires, les ADNc marqués s'hybrident aux sondes complémentaires fixées sur le support. Une fois le support lavé, une image de la puce est réalisée. La fluorescence ou la radioactivité de chaque point de la puce est détectée et convertie en une intensité lumineuse par un

scanner approprié. Une analyse bioinformatique permet alors de comparer les quantités relatives de chaque espèce d'ARNm au sein des échantillons.

Le séquençage de l'ADN a été mis au point dans les années 1970 indépendamment par Frederick Sanger (Royaume-Uni) et Walter Gilbert (Etats-Unis). Ces deux scientifiques ont reçu en 1980 le prix Nobel de chimie. La méthode de Gilbert ayant recours à des réactifs très toxiques, elle a été progressivement abandonnée au profit de la méthode de Sanger. Cette méthode a connu d'énormes progrès technologiques afin de permettre le séquençage de manière parallèle et à bas coût de génomes eucaryotes entier en peu de temps. La possibilité d'obtenir pour un coût modéré et dans un temps raisonnable les séquences d'un ensemble d'acides nucléiques de grande taille a conduit à transposer cette technique à l'analyse du transcriptome (RNA-seq, figure 7) [21]. La première étape, commune aux autres techniques, est l'extraction de l'ARN. Une des difficultés est d'éliminer les espèces trop abondantes qui risquent de saturer l'analyse, comme par exemple les ARN ribosomiques. Les ARNm sont alors convertis en ADNc. Différents procédés spécifiques [22-27] ont été développés pour retrouver le brin à partir duquel l'ARN a été transcrit, comme par exemple la synthèse d’ADNc mono-brin (Illumina [28]). Le tableau 1 présente les trois principaux systèmes de séquençage haut-débit avec leurs caractéristiques. La phase d'hybridation est virtuelle, par alignement des séquences obtenues avec un génome de référence. Une fois les séquences alignées sur le génome, elles sont dénombrées, ce qui permet une analyse quantitative au niveau de chaque gène.

Réponse à l'infection : données expérimentales

Réponse cellulaire à l'infection

Les puces à ADN permettent *(1)* d'identifier des signatures moléculaires spécifiques en fonction du stimulus (cellules en culture stimulées par de l'interféron (IFN) α, β ou γ) [30] ; *(2)* de décrire la réponse de l'hôte à l'infection [31]. En utilisant des cellules dendritiques, impliquées dans l'initiation des réponses immunes innée et acquise, des profils de réponse spécifique et commune à différents micro-organismes ont été caractérisés [32]. Ce résultat a été confirmé par l'analyse de la réponse transcriptionnelle de macrophages humains stimulés *ex-vivo* par diverses bactéries (*E. coli, Salmonella sp., Staphylococcus aureus, Listeria monocytogenes, Mycobacterium tuberculosis,* et le bacille de Calmette et Guerrin) [33], et sur sang total, mélangé *ex-vivo* à du lipopolysaccharide (LPS) ou des fragments de *Staphylococcus aureus* [34]. La réponse de l'hôte au niveau cellulaire a donc été caractérisée au sein de plus de 25 types cellulaires, après stimulation par des bactéries, virus, levures ou parasites [35]. Ainsi, l'analyse de la réponse transcriptionnelle de l'hôte est un outil de diagnostic moléculaire caractérisant au sein de diverses cellules un profil de réponse temporel précis, sensible et spécifique pour divers micro-organismes.

Spécificité de la réponse (selon le tissu ou le micro-organisme)

Plusieurs équipes ont essayé de confirmer ces résultats *in-vivo*, dans des modèles animaux de sepsis. L'analyse *in-vivo* introduit un degré supplémentaire de complexité puisque la réponse de l'hôte peut être observée au sein d'un ou plusieurs tissus. Dans

un modèle de sepsis intra-abdominal plurimicrobien (secondaire à une perforation et ligature caecale [CLP]), la réponse tissulaire (analysée au niveau du cerveau, des reins, du foie, des poumons, de la rate et du thymus) était majoritairement spécifique de l'organe, parfois commune à plusieurs organes, ou nettement opposée entre différents organes [36]. L'analyse de la littérature montre que les profils d'expression des gènes diffèrent dans le sang, le tissu hépatique et le tissu splénique [37,38]. La réponse de chaque organe a ainsi été analysée de façon spécifique : cœur [39-42], cerveau [36,43,44], foie [36,37,45-49], rate [36-38,46,47,50,51], poumon [36,47], et sang [38,52,53]. La variabilité des modèles utilisés rend difficile la synthèse de ces résultats. Il apparaît simplement qu'il existe une réponse tissulaire spécifique.

L'utilisation d'un stimulus différent (CLP (polymicrobien) ou LPS) est également à l'origine d'une variation de la réponse [38]. L'analyse du transcriptome aux niveaux sanguin et splénique identifie des signatures spécifiques de chaque stimulus (CLP ou LPS). Dans une autre étude, l'analyse du transcriptome hépatique montre que l'expression de 17 gènes diffère après infection par *Escherichia coli* ou *Staphylococcus aureus* [49]. La réponse est donc spécifique, selon l'organe analysé ou la nature de la stimulation.

Pronostic

L'analyse du transcriptome est une aide pour classer les états septiques selon leur gravité. Peu d'études ont exploré cette question chez l'animal. Expérimentalement, les profils d'expression discriminent avec précision des groupes de souris exposées à *(1)* une injection de sérum physiologique (contrôles), *(2)* l'administration de LPS, *(3)*

une laparotomie blanche, *(4)* une laparotomie associée à une CLP, et *(5)* une CLP associée à un traitement antibiotique [38]. Les différents groupes de souris avaient des taux de mortalités variant de 0 % à 100 %. En pathologie humaine, on envisage donc que l'analyse du transcriptome soit une aide pour identifier les patients à risque de décès.

Distinction SIRS/Sepsis

Au lit du malade, déterminer l'origine infectieuse ou non des états inflammatoires reste difficile. L'implication thérapeutique est directe puisque la prescription d'antibiotiques repose sur le résultat de cette analyse. Demain, des traitements immunomodulateurs pourraient être prescrits en fonction de cette différence. Des signatures spécifiques ont été mises en évidence en comparant le transcriptome exprimé par des PBMC prélevés chez des rats « contrôles » (sérum physiologique), des rats ayant une infection intra-abdominale (CLP) et des rats présentant une réponse inflammatoire systématisée (pancréatite) (Figure 8) [53]. Seuls 15 (5 %) gènes modulés sont communs aux groupes CLP et pancréatite. De la même manière, il existe une signature spécifique selon que des souris sont soumises à une laparotomie blanche ou associée à une CLP [38]. Ces résultats soutiennent l'hypothèse selon laquelle l'analyse du transcriptome permet de distinguer SIRS et sepsis.

Intégration du facteur temps

L'analyse du transcriptome a permis de caractériser l'évolution temporelle des profils d'expression génique durant un épisode septique. Dans un modèle murin, le

transcriptome hépatique a été analysé à 2 h, 8 h, 24 h 48 h et 72 h après injection intraveineuse d'*Escherichia coli* ou de *Staphylococcus aureus* [49]. Sur cette période, 300 (5 %) gènes étaient modulés par un facteur 2 ou plus. Ces gènes étaient regroupés en quatre groupes fonctionnels, *(1)* des gènes de la réponse inflammatoire et de la réponse au stress, *(2)* des gènes codant des protéines de structure, et des gènes impliqués dans la motilité cellulaire ainsi que dans le trafic vésiculaire au sein de la cellule, *(3)* des gènes codant des récepteurs membranaires, des molécules de transduction du signal et des facteurs de transcription, *(4)* des gènes codant des enzymes du métabolisme cellulaire, ainsi que des gènes impliqués dans la synthèse protéique. L'expression des gènes au cours du temps évoluait selon quatre profils : *(1)* augmentation ou *(2)* diminution de l'expression, suivie d'un retour à l'état basal ; *(3)* augmentation ou *(4)* inhibition persistante de l'expression.

Dissection du modèle physiopathologique chez l'animal

Un des attraits des modèles animaux est de comparer la modulation de l'expression des gènes chez des animaux sauvages ou génétiquement modifiés (KO pour « Knock-Out » en anglais). Le rôle de certains gènes dans la réponse au sepsis peut ainsi être précisé. En étudiant un modèle de CLP au sein de souris sauvages ou ayant une délétion du gène codant MyD88 (MyD88-/-), on observe une diminution du nombre de gènes modulés au niveau hépatique chez les souris KO MyD88-/-. Ce résultat n’est pas retrouvé au niveau splénique. Cela confirme le rôle de certains gènes dans la réponse tissulaire spécifique. L’analyse du transcriptome hépatique de souris IL6ST KO montre qu’IL6ST (également nommée GP130) est impliqué dans la

transduction du signal de cytokines comme l'IL-6 vers STAT3. Surtout, ce gène est essentiel pour la survie des souris, via l'activation de la réponse inflammatoire et le recrutement de cellules immunitaires au niveau hépatique [48]. Il a également été montré que la NO synthase inductible est essentielle pour moduler un certain nombre de gènes associés à la dysfonction myocardique dans le sepsis [41].

Intégration de la variabilité inter-individuelle

L'outil d'analyse du transcriptome permet d'explorer la variabilité inter-individuelle en réponse à l'infection. Celle-ci peut être d'origine génétique ou environnementale. En analysant les réponses hépatique, splénique et pulmonaire de différentes lignées de souris à l'injection de LPS, de nouveaux facteurs de transcription impliqués dans le contrôle de l'expression des gènes induits par le LPS ont été identifiés [47]. Par rapport aux souris sauvages, les souris ayant une délétion du gène codant pour le facteur de transcription E2F1 avaient une réponse inflammatoire réduite après l'administration systémique de LPS. Les souris hétérozygotes pour le gène PTCH, impliqué dans la voie d'Hedgehog, avaient au contraire une réponse exacerbée au LPS. Enfin, des facteurs intrinsèques, comme l'âge ou l'obésité, modifient également les profils de réponse transcriptionnels [40,43,54].

L'ensemble de ces études répond à différentes questions : *(1)* la description de la réponse transcriptionnelle à un micro-organisme au cours du temps ; *(2)* la description des réponses commmunes et spécifiques de l'hôte selon le tissu ou le

stimulus ; *(3)* l'altération de cette réponse lors de l'utilisation d'un hôte mutant (pour décrire le rôle de certains gènes dans la réponse, et ainsi identifier des cibles thérapeutiques potentielles). Ces différents travaux chez l'animal ont posé les bases de l'analyse du transcriptome en pathologie humaine. Ils permettent également d'affiner la compréhension des mécanismes physiopathologique du sepsis.

Réponse à l'infection : données humaines

L'ensemble des résultats obtenus *in vitro* et *in vivo* ont conduit à utiliser les puces à ADN comme un outil d'analyse du transcriptome chez les patients septiques. Les objectifs étaient *(1)* d'identifier une réponse transcriptionnelle liée à l'infection ; *(2)* de déterminer si comme chez l'animal, la réponse est spécifique de chaque organe tissus ; *(3)* de déterminer si cette réponse distingue les états inflammatoires (SIRS) d'origine infectieuse ou non ; *(4)* de déterminer si chaque micro-organisme a une signature spécifique; *(5)* de déterminer si une signature temporelle permet de suivre l'évolution de la maladie ; et *(6)* de déterminer si des profils spécifiques sont liés au pronostic.

Réponse à l'infection

L'analyse du transcriptome à partir de PBMC ou de sang total identifie les sujets septiques. A l'aide d'une puce à ADN (testant 340 gènes humains impliqués dans la réponse inflammatoire), le transcriptome sanguin (sang total) a été comparé chez huit patients en sepsis grave et quatre patients en surveillance post-opératoire d'une

chirurgie du rachis, sans signe de SIRS [55]. Cette analyse différenciait les deux groupes de patients. A l'aide d'une puce Affymetrix (GeneChip, 54 000 sondes), une signature distinguant des patients septiques de sujets contrôles a également été identifiée en réanimation pédiatrique [56]. Elle comprenait des gènes impliqués dans la réponse immunitaire et dans l'homéostasie du zinc. Dans une étude de confirmation, 95 à 100 % des enfants étaient classés dans la bonne catégorie diagnostique [57].

Réponse tissulaire à l'infection

Une seule étude humaine a analysé le transcriptome de cellules musculaires striées dans le sepsis. L'objectif était de décortiquer les phénomènes de myopathie liée aux états septiques graves [58]. Toutefois, il n'existe pas de données chez l'homme comparant les réponses transcriptionnelles des différents tissus dans le sepsis.

Distinction des états inflammatoires d'origine infectieuse ou non

Du point de vue du clinicien, un des enjeux est de distinguer l'origine infectieuse ou non d'un syndrome inflammatoire. En réanimation pédiatrique, il existe des différences entre le transcriptome de cellules mononuclées d'enfants en SIRS, sepsis, et choc septique [59]. Une partie de la modulation de l'expression des gènes est commune à l'ensemble des patients. Les auteurs l'attribuent à la réponse inflammatoire. Ils identifient également une réponse spécifique des patients septiques par rapport aux patients inflammatoires. L'analyse temporelle montre que la modulation des gènes persiste trois jours seulement chez les enfants en choc septique.

Cette modulation est majoritairement une sous-expression de gènes liés à la réponse immunitaire acquise et à l'homéostasie du zinc.

Des résultats similaires ont été retrouvés chez l'adulte. Le transcriptome de cellules mononuclées a été étudié chez 46 patients en sepsis et 24 patients de réanimation avec un diagnostic de syndrome de réponse inflammatoire systémique non liée à l'infection [60]. La sensibilité et la spécificité de cette signature ont été testées sur un second groupe de patients. Elles étaient de 81 % et de 79 %, respectivement [60]. L'évolution de l'expression des gènes sur trois jours a été comparée chez des patients en SIRS évoluant ou non vers un sepsis. L'expression de 4 559 gènes discrimine les deux groupes de patients [61]. Globalement, les gènes identifiés sont impliqués dans la réponse immune innée, la transduction du signal des cytokines, la différentiation des lymphocytes T et la régulation de la synthèse protéique. Alors que les gènes impliqués dans la réponse inflammatoire sont fortement activés dans les états de SIRS, ils sont sous-exprimés dans les états septiques graves. Par contre, les gènes impliqués dans l'apoptose sont activés. Ces résultats soutiennent la vision actuelle de la physiopathologie du sepsis grave avec une réponse anti-inflammatoire exagérée et un état d'anergie.

Identifier une réponse spécifique d'un micro-organisme

Une seule étude animale a montré des différences au sein du transcriptome selon le micro-organisme en cause [49]. Les tentatives pour reproduire ces résultats en pathologie humaine sont infructueuses. Ceci est bien corrélé à une étude clinique montrant une absence d'association entre la nature de la bactérie et l'évolution du

patient [62]. Le transcriptome de 131 enfants atteints d'infection aiguë a été analysé en fonction de la nature virale ou bactérienne de l'infection. Trente-cinq gènes distinguaient avec précision l'infection en fonction de la nature de l'agent microbien [63]. Toutefois, aucune différence n'était notée entre les différentes bactéries. Deux études conduites sur des populations d'adultes hospitalisés en réanimation n'ont pas eu la capacité de discriminer une infection par une bactérie à Gram positif ou négatif [60,64]. La contradiction entre les résultats animaux et humains est probablement liée à un manque de puissance de l'analyse. En effet, dans le modèle expérimental, les animaux ont un fonds génétique commun, et les prélèvements sont réalisés dans un même délai par rapport à l'inoculation. De plus, seules deux bactéries distinctes étaient utilisées et un faible nombre de gènes identifiés (n = 17) [49]. Chez l'homme, il existe une variabilité du « fonds génétique », une variabilité en terme de délai de prélèvement par rapport au début de l'infection, et une variabilité de micro-organismes en cause, avec parfois des infections polymicrobiennes. Enfin, la majeure partie de la réponse transcriptionnelle est probablement commune. Il est donc possible de caractériser les mécanismes moléculaires spécifiques mis en jeu par divers agents pathogènes dans les modèles animaux, mais il est actuellement difficile d'utiliser les données de transcriptome pour affiner le diagnostic microbiologique, en dehors de la distinction entre infections virale et bactérienne.

Evolution naturelle ou en réponse aux traitements de la réponse transcriptionnelle

Plusieurs études ont décrit une signature chez des sujets sains exposés à une

injection intraveineuse de LPS. Le LPS est un des composants de la membrane externe des bactéries à Gram négatif. Il joue un rôle prépondérant dans le déclenchement de la réponse immunitaire lors de l'infection [65-67]. Après administration de LPS, l'expression d'environ 3 700 gènes est modulée avec un retour à l'expression basale dans les 24 h. La plupart des gènes sont activés entre 4 et 6 h après l'agression. La transcription de gènes impliqués dans le rétrocontrôle de l'activation de la réponse immunitaire s'articule durant cette période. Parmi les gènes modulés, on note une sur-expression des gènes impliqués dans le système de phagocytose NADPH-oxydase. Ce composant est important dans le système de défense de l'hôte. On note également une diminution de l'expression des gènes impliqués dans le complexe majeur d'histocompatibilité de type II et une réduction de la synthèse des intégrines, impliquées dans l'adhésion intercellulaire (figure 9).

Une analyse longitudinale du transcriptome permettrait de suivre l'évolution naturelle de la maladie et la réponse au traitement. L'expression de certains biomarqueurs impliqués dans la réponse pro-inflammatoire (S100A8, S100A12, ALOX5AP) diminue au cours de la rémission du choc septique, alors que l'expression de gènes impliqués dans la présentation de l'antigène (par exemple CD74) augmente [68]. L'évolution du transcriptome a été comparée chez 90 patients en SIRS pendant les 72 h précédant ou non l'apparition du sepsis [69]. L'analyse, conduite *a posteriori,* s'est focalisée sur les gènes impliqués dans la transduction du signal par les récepteurs de type Toll. L'expression de plusieurs de ces gènes est discriminante un à deux jours avant la survenue des signes cliniques. McDunn et al.

ont proposé à partir de données de transcriptome le concept de riboleukogramme (figure 10) [70]. Une signature transcriptionnelle a été déterminée par l'analyse des modifications du transcriptome obtenues dans un modèle murin de pneumonie. Cette signature a ensuite été analysée chez des patients dont les poumons sont ventilés mécaniquement. L'expression des 85 gènes de la signature permet de suivre l'évolution d'un épisode de pneumonie, avec un diagnostic moléculaire précédant le diagnostic clinique, et une seconde modulation du riboleukogramme lors de la rémission. Malheureusement, ce riboleukogramme n'a pas été validé dans une population de 158 patients traumatisés graves [71]. La sensibilité et la spécificité diagnostique du test étaient respectivement de 57 % et 69 %. Toutefois, en dépit de ce résultat, un tel modèle d'analyse reste probablement pertinent pour le diagnostic et le suivi de la réponse au traitement.

Identification de signatures pronostiques

En clinique, il existe enfin un grand intérêt pour la recherche de biomarqueurs pronostiques. L'analyse du transcriptome s'intègre totalement dans cette démarche. Des signatures spécifiques ont été associées à un mauvais pronostic. Chez des patients en choc septique, une signature de 28 gènes classe les patients selon leur survie avec une sensibilité de 100 % et une spécificité de 86 % [72]. De la même manière, Wong *et al.* ont identifié 34 gènes dont l'expression prédit la mortalité d'enfants en choc septique [73]. De façon surprenante, la comparaison des deux signatures ne retrouve aucun gène commun. Il est probable que les résultats obtenus ne soient valides qu'au sein de la population étudiée (enfants *vs.* adultes). Les

signatures identifiées sur ces analyses pronostiques diffèrent également de celles identifiées pour la détection du sepsis [74].

L'ensemble de ces travaux souligne la variabilité importante de la réponse de l'hôte à l'infection. Parmi les variables qui influencent la réponse de l'hôte, le sexe a été particulièrement étudié dans le sepsis. Plusieurs travaux montre que le sexe modifie la réponse à l'infection et le pronostic des patients en sepsis [75,76]. Pourtant, bien qu'il semble évident que le sexe joue un rôle dans la réponse transcriptionnelle, les divers travaux d'analyse globale du transcriptome par puces à ADN n'ont pas pris en compte cette variable dans leur analyse.

Au total, l'analyse du transcriptome par la technologie des puces à ADN permet de décrire la réponse de différents tissus à l'agression, et d'affiner notre compréhension des mécanismes physiopathologique de l'inflammation et du sepsis. C'est donc un outil d'analyse de la réponse de l'hôte à l'agression, notamment à l'infection.

Objectifs

L'objectif de cette thèse est d'explorer l'inflammation et l'infection au niveau du transcriptome, à l'aide de la technologie des puces à ADN.

Dans ce but, nous avons mis au point un logiciel permettant d'explorer et d'analyser les données publiques de transcriptome. Ce travail a été initié devant le constat d'une croissance exponentielle des données publiques de transcriptome, et la pauvreté d'outils adaptés pour comparer et analyser de manière intégrative ces données et les autres bases de données publiques d'annotation.

Ensuite, nous avons essayé de différencier la réponse inflammatoire de l'hôte, associée ou non à un sepsis. La volonté de distinguer une réponse infectieuse spécifique au sein de la réponse inflammatoire nous a conduit à mettre en place un schéma expérimental permettant de comparer une réponse inflammatoire liée ou non à l'infection. Pour cela, nous avons mis au point un modèle murin d'atteinte inflammatoire pure du poumon par injection d'acide oléique et étudié dans un premier temps la réponse du parenchyme pulmonaire, et des cellules sanguines mononuclées. Ainsi, il sera possible de comparer dans un second temps la réponse de ce modèle d'atteinte inflammatoire, à un modèle murin d'infection pulmonaire. Nous avons ensuite imaginé un modèle clinique similaire, analysant des traumatisés graves dont l'évolution est compliquée d'une pneumonie associée à la ventilation mécanique (projet Sepsi*Chip*). Chez un même patient, le syndrome inflammatoire déclenché par le traumatisme grave sert ainsi de témoin à la réponse infectieuse de la pneumonie.

Le second objectif de la thèse est d'utiliser l'analyse globale du transcriptome pour explorer les mécanismes physiopathologiques du sepsis, et notamment les variables qui influencent la relation hôte-pathogène, et la réponse de l'hôte à l'agression. Les deux premiers modèles d'analyse (projet Sepsi*Chip* chez l'homme, et le modèle murin d'atteinte inflammatoire pulmonaire) permettent de souligner le rôle du temps.

Comme le sexe est une des principales caractéristiques de l'hôte et que de nombreux travaux soutiennent son rôle dans la réponse à l'agression, nous avons choisit d'intégrer ce facteur dans l'analyse. Le choix d'un modèle murin d'infection par *Coxiella burnetii* repose sur l'existence de travaux antérieurs du laboratoire/de l'équipe sur le rôle du sexe dans la fièvre Q, et l'existence d'un dimorphisme sexuel connu en clinique. Enfin, pour intégrer dans l'analyse une variable liée au pathogène, le dernier modèle analysé (infection par le virus *Influenza*) examine l'influence de l'infection par différentes souches plus ou moins virulentes d'un même virus sur la réponse de l'hôte (dans un modèle de cellules en culture).

Le dernier objectif de ce travail est d'évaluer si l'analyse globale de la réponse de l'hôte peut conduire à l'identification de nouvelles cibles thérapeutiques, au sein d'un modèle *in-vitro* d'infection virale par différentes souches du virus de la grippe. Au cours de ce travail, nous avons mis au point une méthodologie originale *in silico* d'identification de molécules candidates. Huit antiviraux potentiels ont ainsi été évalués *in-vitro* dans le même modèle d'analyse.

Résultats

TranscriptomeBrowser: A Powerful and Flexible Toolbox to Explore Productively the Transcriptional Landscape of the Gene Expression Omnibus Database

Julien Textoris, Fabrice Lopez, Aurélie Bergon, Gilles Didier, Elisabeth Remy, Samuel Granjeaud, Jean Imbert, Catherine Nguyen, Denis Puthier

PLoS ONE 3(12): e4001

L'utilisation de technologies à haut-débit comme par exemple les puces à ADN génère une quantité importante de données. Ces données sont aujourd'hui accessibles publiquement dans des bases de données en ligne. La taille de ces bases grandit rapidement, et nécessite la mise au point de stratégies d'analyse pour traiter ces données de manière efficace.

Nous avons utilisé une version modifiée de l'algorithme MCL (pour Markov CLustering) pour extraire de manière systématique et automatisée des clusters de gènes co-régulés à partir de centaines de jeux de données (n = 1 484) stockés dans la base de donnée GEO (pour « Gene Expression Omnibus », NCBI). Cette stratégie a conduit à l'identification de 18 250 signatures transcriptionnelles. Ces signatures ont dans un second temps été annotées à partir de la suite logicielle DAVID (pour « Database for Annotation, Visualization and Integrated Discovery »). Un enrichissement fonctionnel a été retrouvé pour la plupart de ces signatures transcriptionnelles (84 %). Nous avons développé une application java modulaire et évolutive, *TBrowser* qui est composée d'une interface graphique et d'une série de greffons. L'architecture open source autorise le développement de nouveaux greffons d'analyse par la communauté scientifique. L'interface graphique est muni d'un module de recherche sophistiqué supportant les opérations booléennes (http://tagc.univ-mrs.fr/tbrowser/). L'utilisateur peut interroger la base de données de signatures transcriptionnelles à partir de divers identifiants (noms de gènes, identifiants de sondes, …) et/ou des mots-clés d'annotation fonctionnelle.

Comme preuve de concept, *TBrowser* a été utilisé pour identifier des gènes spécifiques de tumeur maligne du sein, ainsi que pour détecter des anomalies chromosomiques associées à ces tumeurs. Enfin, à partir de cette vaste collection de signatures transcriptionnellles, nous avons construit une carte résumant les co-régulations gène/gène observées à travers l'ensemble des expériences réalisées sur une puce à ADN Affymetrix HGU133A. Nous montrons ainsi que l'analyse de cette carte permet de compléter nos connaissances sur diverses voies métaboliques. Enfin, dans l'analyse de la réponse à l'infection par *Coxiella burnetii* en fonction du sexe, l'utilisation de *TBrowser* nous a permis de mettre en évidence des différences de localisation sub-cellulaire.

OPEN ACCESS Freely available online

PLoS one

TranscriptomeBrowser: A Powerful and Flexible Toolbox to Explore Productively the Transcriptional Landscape of the Gene Expression Omnibus Database

Fabrice Lopez[1,2☯], Julien Textoris[1,2,5☯], Aurélie Bergon[1,2], Gilles Didier[2,3], Elisabeth Remy[2,3], Samuel Granjeaud[1,2], Jean Imbert[1,2], Catherine Nguyen[1,2], Denis Puthier[1,4]*

1 Inserm U928, TAGC, Parc Scientifique de Luminy, Marseille, France, **2** Université de la Méditerranée, Marseille, France, **3** Institut de Mathématiques de Luminy, Campus de Luminy, Marseille, France, **4** ESIL, Université de Provence et de la Méditerranée, Marseille, France, **5** Service d'Anesthésie et de Réanimation, hôpital Nord - Assistance Publique, Hôpitaux de Marseille, Marseille, France

Abstract

Background: As public microarray repositories are constantly growing, we are facing the challenge of designing strategies to provide productive access to the available data.

Methodology: We used a modified version of the Markov clustering algorithm to systematically extract clusters of co-regulated genes from hundreds of microarray datasets stored in the Gene Expression Omnibus database (n = 1,484). This approach led to the definition of 18,250 transcriptional signatures (TS) that were tested for functional enrichment using the DAVID knowledgebase. Over-representation of functional terms was found in a large proportion of these TS (84%). We developed a JAVA application, TBrowser that comes with an open plug-in architecture and whose interface implements a highly sophisticated search engine supporting several Boolean operators (http://tagc.univ-mrs.fr/tbrowser/). User can search and analyze TS containing a list of identifiers (gene symbols or AffyIDs) or associated with a set of functional terms.

Conclusions/Significance: As proof of principle, TBrowser was used to define breast cancer cell specific genes and to detect chromosomal abnormalities in tumors. Finally, taking advantage of our large collection of transcriptional signatures, we constructed a comprehensive map that summarizes gene-gene co-regulations observed through all the experiments performed on HGU133A Affymetrix platform. We provide evidences that this map can extend our knowledge of cellular signaling pathways.

Citation: Lopez F, Textoris J, Bergon A, Didier G, Remy E, et al. (2008) TranscriptomeBrowser: A Powerful and Flexible Toolbox to Explore Productively the Transcriptional Landscape of the Gene Expression Omnibus Database. PLoS ONE 3(12): e4001. doi:10.1371/journal.pone.0004001

Editor: Pamela A. Silver, Harvard Medical School, United States of America

Received July 8, 2008; **Accepted** November 25, 2008; **Published** December 23, 2008

Funding: This work was supported by the Institut National de la Santé et de la Recherche Médicale (Inserm), the Cancéropçle PACA and Marseille-Nice Genopole®. Fabrice Lopez was supported by a fellowship from the EU STREP grant Diamonds and through funding from the IntegraTCell project(ANR, National Research Agency). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: puthier@tagc.univ-mrs.fr

☯ These authors contributed equally to this work.

Introduction

Microarray technology provides biologists with a powerful approach for comprehensive analyzes of cells or tissues at the transcriptional level. DNA chips are now widely used to assess the expression levels from all genes of a given organism. These data, most generally deposited in MIAME-compliant public databases, constitute an unprecedented source of knowledge for biologists [1]. As an example, until now, the Gene Expression Omnibus repository (GEO) host approximately 8,000 experiments encompassing about 200,000 biological samples analyzed using various high through-put technologies [2]. Consequently, this represents billions of measurements that reflect the biological states of cells or tissues recorded in physiological or pathological conditions or in response to various chemical compounds and/or natural molecules. As public repositories are continually expanding, we are facing the new challenge of designing new strategies to provide efficient and productive access to the available data.

To date, at least two major solutions have emerged. The first one applies a "*gene-centered perspective*", as developed in the "GEO profile" or "SOURCE" web interfaces [3]. This approach allows users to retrieve the expression profiles of a given gene in numerous curated experiments. Once a profile is selected, a list of similar profiles (*i.e.* neighbors) can be retrieved. Although GEO proposes several tools to refine the queries, cross-analysis through multiple experiments can not be performed. The second solution involves an "*experiment-centered perspective*" as developed in the "GEO DataSets" and "ArrayExpress" web interfaces [4]. This approach provides to biologists a set of classification tools to re-analyze selected experiments. Depending on the interface, supervised or unsupervised analysis (see below) can be pre-calculated or computed on demand. Again, as no meta-analysis tool is available, mining and compiling even few GEO Serie Experiments (GSE) remains a difficult and time-consuming task.

We therefore lack efficient tools allowing productive data mining of microarray databases. For example, querying whole

public microarray data using a single gene identifier is an ambiguous procedure to extract relevant co-regulated genes. Indeed, depending of the biological context, genes can be involved in different signaling pathways and may be associated with different neighbors. As a consequence, combined queries should be more appropriate to build relevant gene networks. Moreover, numerous uninformative genes exist in microarray experiments. They correspond most generally to those with low standard deviation that are outside any natural gene cluster. These genes should be discarded from analysis as they are inevitably associated with false positive neighbors. These considerations motivated the present work and the development of a new approach that follows a "*transcriptional signature centered perspective*". The goal was to build an application that would interact with a large database of transcriptional signatures and would implement efficient tools to analyze and visualize the results.

The first issue resided in the construction of a database containing high quality transcriptional signatures obtained in an automated fashion. Both supervised and unsupervised classification algorithms can be used in microarray data analysis [5]. Supervised methods aim at finding a set of genes whose expression profiles best correlate with a known phenotype. They provide a way to select informative genes by choosing the top k genes according to the results of a statistical test (*e.g.* Student's *t*-test, Significance Analysis of Microarrays, Signal to Noise Ratio, ANOVA) and by controlling the false discovery rate (FDR). In contrast, unsupervised classification approaches, achieve clustering of genes based on their respective expression profiles but are not intended to filter out uninformative genes. Some popular approaches in microarray analysis use either agglomerative methods (hierarchical clustering), partitioning methods (k-medoids, k-means, PAM, SOM, etc.) or methods aimed at capturing informative dimensions (PCA). A filtering step is most generally applied prior to unsupervised classification. One can select genes with high standard deviations, those displaying a proportion of values above a user-defined threshold or those having a given maximum (or minimum) value. However this procedure is extremely subjective and the number of selected genes may be over or under estimated. Finally, another limit of classical unsupervised methods also resides in their inability to accurately identify the actual number of clusters if no further argument is provided to the algorithm. As a consequence, additional algorithms for unsupervised classification have been proposed such as Quality Cluster algorithm (QT_Clust) [6], CHAMELEON [7] or Markov CLustering (MCL) [8]. However, none of them address both the filtering and partitioning issues. MCL is a graph partitioning algorithm whose ability to solve complex classification problems has been underlined in many applications including protein-protein interaction networks [9], sequence analysis (TRIBE-MCL) [10] or microarray analysis (geneMCL) [11]. In a graph representation of microarray data, nodes stand for genes and edges represent profile similarities between genes. As processing the full graph for partitioning is time-consuming and computer-intensive the geneMCL algorithm has to be run on a subset of genes that are selected using classical filters (*e.g.* high standard deviation or fold-change). As such a filtering procedure is not well suited for automated analysis of numerous experiments; we developed an adaptive density-based filter (DBF) whose goal is to isolate automatically informative genes from a dataset. Selected genes are next used to construct a graph that is subsequently partitioned using MCL. This modified version of MCL algorithm was termed DBF-MCL for "Density Based Filtering and Markov CLustering".

In the present paper, we show that DBF-MCL provides very good results both on simulated and real datasets. The algorithm was run on 1,484 microarrays datasets (46,564 biological samples) performed on various Affymetrix platforms (human, mouse and rat). This led to the identification of 18,250 transcriptional signatures (TS) whose corresponding gene lists were tested for an enrichment in terms derived from numerous ontologies or curated databases using the DAVID knowledgebase [12] (Gene Ontology, KEGG, BioCarta, Swiss-Prot, BBID, SMART, NIH Genetic Association DB, COG/KOG, etc.) (see Figure S1 for an overview of the data processing pipeline). Informations related to biological samples, experiments, TS composition, TS associated expression values and TS keyword enrichment scores were stored in a relational database. A Java application, TBrowser (TranscriptomeBrowser), was developed and deployed using Java Web Start technology. Combined queries that can be done with an extended set of Boolean operators allow user to rapidly select sets of TS containing (or not) a given list of gene symbols. Based on these TS, a list of frequently observed neighbors can be created. As each TS is linked to a set of biological keywords (derived from ontologies), user can also search for those enriched in genes involved in specific biological processes. We show that TBrowser can be used to mine productively hundreds of experiments and to reveal underlying gene networks. Furthermore, using this unprecedented collection of TS we built the first synthetic transcriptional map of all human microarray data performed on Affymetrix HG-U133A platform and currently available in the GEO database.

Results

DBF-MCL algorithm

Conventional algorithms used for unsupervised classifications of gene expression profiles suffer from two main limitations. First, they do not filter out uninformative profiles and second, they are not able to find out the actual number of natural clusters in a microarray dataset. We can considerer genes as points located in a hyperspace whose number of axes would be equal to the number of biological samples. As it is difficult to perceive high-dimensional spaces, a common way to illustrate classification methods is to use a 2D representation. In Supplemental Figure S2, each point represents a gene and we are interested in isolating dense regions as they are populated with genes that display weak distances to their nearest neighbors (*i.e.* strong profile similarities). To isolate these regions we can compute, for each gene, the distance with its k_{th} nearest neighbor (DKNN). If k is relatively small, DKNN should be smaller for all genes falling in a dense area. Thus, the filtering procedure used in DBF-MCL starts by computing a gene-gene distance matrix D. Then, for each gene, DBF-MCL computes its associated DKNN value (with k being set typically to 100 for microarrays containing 10 to 50k elements). Distributions of DKNN values observed with both an artificial and a real dataset (Complex9RN200 and GSE1456 respectively, see thereafter for a description) are shown in Figure S3A and S3B (solid curve). The asymmetrical shape of the distribution observed in Figure S3B suggests the presence of a particular structure within the GSE1456 microarray dataset. Indeed, the long tail that corresponds to low DKNN values could indicate the existence of dense regions. The fact that regions of heterogeneous densities exist in the Complex9RN200 artificial dataset is even clearer as a bimodal distribution is observed. Next, we would like to define a critical DKNN value below which a gene can be considered as belonging to a dense area and that would depend on the intrinsic structure of the dataset. To this end, DBF-MCL computes simulated DKNN values by using an empirical randomization procedure. Given a dataset containing n genes and p samples, a simulated DKNN value is obtained by sampling n distance values from the gene-gene distance matrix D and by extracting the k_{th}-smallest value. This procedure is repeated n times

to obtain a set of simulated DKNN values S_i. As shown in Figure S3 (dotted line), the variance of the simulated DKNN values is very low compare to that observed using the real dataset. Indeed, we can think of simulated DKNN values as the distances to the k_{th} element if no structure existed in the associated space. In this case, we would expect elements to be uniformly spread throughout the space and the variance of DKNN value to be low. In practice several sets $S_{1..q}$ are computed and thus several distributions of simulated DKNN values are obtained. For each observed DKNN value d, a false discovery rate (FDR) value is estimated by dividing the mean number of simulated DKNN below d by the number of observed value below d. The critical value of DKNN is the one for which a user-defined FDR value (typically 10%) is observed. Given a set of selected genes, the next issue is to partition them into homogeneous clusters. This step is achieved through a graph partitioning procedure. In the created graph, edges are constructed between two genes (nodes) if one of them belongs to the k nearest neighbor of the other. Edges are weighted based on the respective coefficient of correlation (*i.e.*; similarity) and the graph obtained is partitioned using the Markov CLustering Algorithm (MCL).

Performances of DBF-MCL on Complex9RN200 dataset

To test the performances of DBF-MCL algorithm we used a modified version of the complex9 dataset which was used earlier by Karypis *et al.* [7]. Since DBF-MCL is designed to handle noisy datasets, 200% of normally distributed random noise was added to the original data. The resulting dataset (which will be referred as Complex9RN200 thereafter, see Figure S4A and S4B) shows some difficulties for partitioning since it is composed of a noisy environment in which arbitrary geometric entities with various spacing have been placed. The two main parameters of DBF-MCL are k that controls the size of the neighborhood and the inflation I (range 1.1 to 5) which controls the way the underlying graph is partitioned. The effect of k on the selection of informative elements is shown in Figure S5A (Euclidean distance was used for this dataset). A steep ascending phase and a slow increasing phase (starting from a k values close to 40) were observed. This confirms the existence of areas with heterogeneous densities. In fact, the transition between the two phases reflects the transition from dense to sparse regions. Indeed, datasets produced with k values above 40 contain noisy elements (Fig S4C). In contrast, choosing k values in the ascending phase ensure the achievement of noise-free datasets. In the case of artificial data, satisfying partitioning results were obtained with inflation values close to 1.2 (Fig S4D–G) although in some cases some of the shapes were merged in a manner that appears to be meaningful (Fig S4E and S4G). We then compared DBF-MCL to several algorithms commonly used in microarray analysis. All of them were run multiple times with various parameters and the best solution was kept. In all cases, the Euclidean distance was used as a distance measure between elements. As these algorithms are not well-suited for noisy data, they were run on the 3,108 points extracted using DBF-MCL (k = 20). Also it is difficult to compare those algorithms to one another, some of them obviously failed to identify the shapes. Indeed, although k-means was run 10 times with random initial starts (and the right number of centers) it led to a very poor partitioning result (Fig S4J). Cluster Affinity Search Technique (CAST, Fig S4K) and Quality Cluster algorithm QT_CLUST (Fig S4I), gave also poor results as did the Self-Organizing Map (SOM) (data not shown). Hierarchical clustering was run with single linkage as arguments and the obtained dendrogram was then split into 9 clusters (Fig S4H). Patterns were well recognized using this method but prior knowledge of the number of clusters is a prerequisite. Thus both DBF-MCL and hierarchical clustering are able to properly identify complex shapes in a 2D space. The main benefit of using DBF-MCL resides in its ability to extract relevant informations from a noisy environment. However, a range of optimal values for inflation parameter needs to be defined to get the best results.

Performances of DBF-MCL on GSE1456 dataset

Next, DBF-MCL was tested with microarray data to explore its effectiveness in finding clusters of co-regulated genes. To this end, we used the microarray data from Pawitan *et al.* [13], who studied gene expression profiles in a large cohort of Swedish patients affected by breast cancer. This experiment is recorded as GSE1456 in the GEO database. All sample (n = 159) have been hybridized onto the GPL96 platform (Affymetrix GeneChip Human Genome U133 Array Set, HG-U133A). The complete dataset (22,283 genes) was used for analysis. Figure S5B, shows the number of informative genes obtained with various k values. Again, two phases were observed suggesting that regions with heterogeneous densities exist in the GSE1456 dataset. As expected, the transition from dense to sparse regions was less marked than in the artificial dataset. A k value of 100 was chosen to allow the extraction of a large part of data that can be considered as noise-free. This value led to the selection of 4,470 elements out of the whole dataset (Fig. 1A–B). The graph partitioning procedure, using default MCL parameters (I = 2), generated 11 highly homogeneous clusters (Fig. 1C–F). As with the Complex9RN200 dataset, the results were very consistent with those obtained using hierarchical clustering although for some genes the clustering results differed (Fig. 1E). Importantly, partition results were not very sensitive to inflation values. Indeed, 10 and 12 clusters were observed with I set to 1.5 and 2.5 respectively (data not shown). All signatures were then submitted to functional enrichment analysis. A summary of the results is given in Figure 1G. As expected for a breast cancer dataset, TS were found to be related to (i) immune response (T-lymphocyte activation, B-lymphocyte activation and interferon alpha), (ii) primary metabolism (cell cycle, ribosome biogenesis, nuclear phosphorylation and transcription) which is probably reminiscent of tumor aggressiveness (iii), modification of local environment (extracellular matrix and cell adhesion) which could sign metastasis potential of each sample, (iv) and estrogen receptor status of breast tumors (estrogen response pathway). Altogether, these results underline the ability of DBF-MCL algorithm to find natural gene clusters within a randomly selected dataset. Indeed, for numerous additional microarray datasets hierarchical clustering results and DBF-MCL results were compared. As illustrated in Figure S5B for a representative set of experiments, setting k to 100 allows in all cases to delete noisy elements and to select only informative genes in a microarray dataset. Interestingly, in all cases meaningful partitioning results were obtained using inflation parameter set to 2.

Systematic extraction of TS

We next applied DBF-MCL algorithm to all experiments performed on human, mouse and rat Affymetrix microarrays and available in the GEO database (33 platforms, Supplementary Table S1 and S2). Only experiments containing more than 10 biological samples were kept for analysis. Overall, this dataset includes 46,564 biological samples hybridized in the context of 1,484 experiments. Each experiment was analyzed independently and subjected to TS discovery process (k = 100, FDR = 10%, $S_{1..3}$, Inflation = 2). As mentioned in the Material and Methods section, we rank-transformed data from each biological sample to get a common input for DBF-MCL algorithm and to allow analysis of a large broad of experiments whose normalization status is frequently unknown. Furthermore, a distance based on Spear-

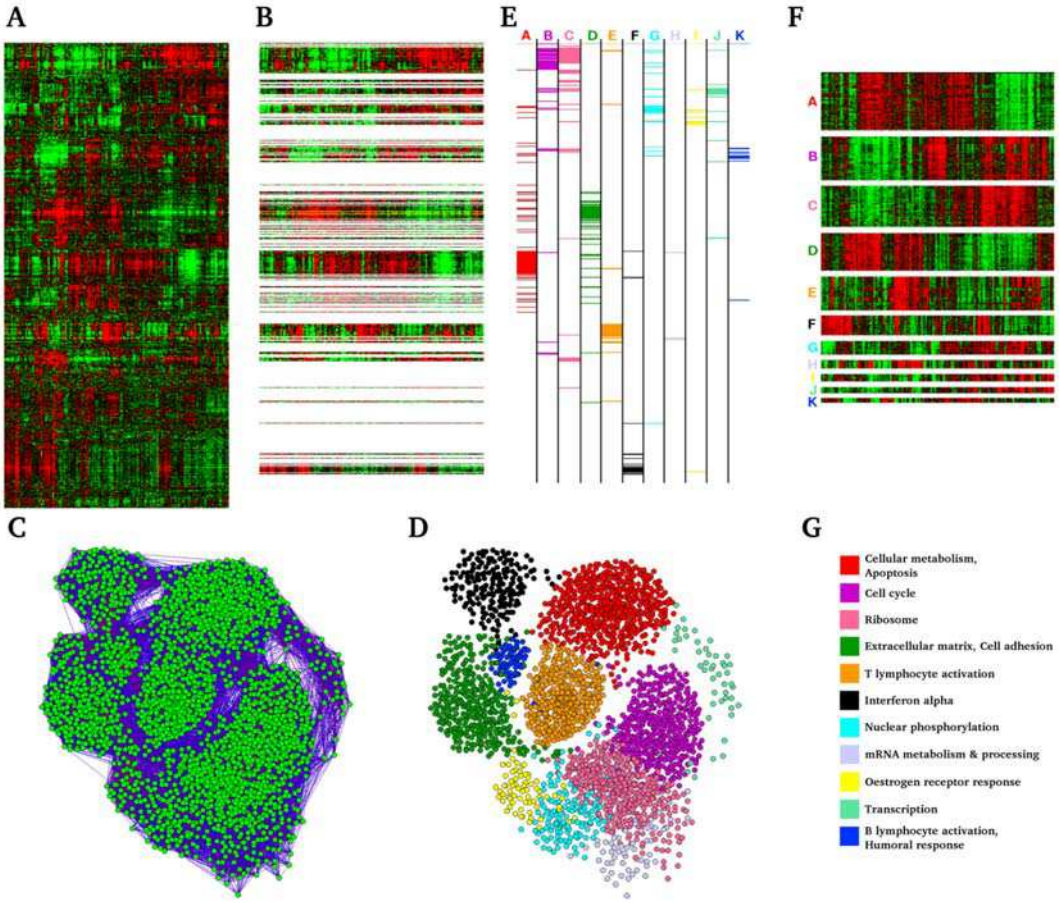


Figure 1. Results obtained with the GSE1456 dataset. DBF-MCL was run with GSE1456 as input (k = 100, FDR = 10%, $S_{1..3}$, Inflation = 2). **(A)** Hierarchical clustering of the GSE1456 dataset. **(B)** Same as (A) but only informative genes are displayed. **(C)** The graph constructed with the 4,470 selected genes. **(D)** The graph after MCL partitioning. Each point is colored according to its associated class. **(E)** Correspondence between hierarchical clustering and DBF-MCL results. **(F)** TS obtained for GSE1456 **(G)** Functional enrichment associated with these TS.
doi:10.1371/journal.pone.0004001.g001

man's rank correlation coefficient was used for k_{th}-nearest neighbor computation. This rank-based distance is known to be clearly more resistant to outlying data points than Pearson-based distance and thus ensured the selection of genes belonging to unmistakable clusters. The full pipeline was run on a server equipped with 6 CPUs and took about 4 days to complete. For the sake of clarity, only results obtained with GPL96 which is the most widely used Affymetrix microarray platform will be presented in this section (311 experiments related to GPL96 were analyzed, 12,752 hybridized samples). On average, 4,341 probes (min = 832, max = 5,849) per expression matrix were declared as informative by DBF-MCL suggesting that routinely 20% of the 22,283 probes measured on the HG-U133A array belong to a natural cluster. Graph partitioning generated on average 10.8 clusters (min = 2, max = 29) for each experiment and each cluster contained approximately 400 probes corresponding in average to 370 distinct gene symbols. Figure 2 shows a summary of these results. As expected no clear correlation was observed between the number of selected genes and the number of samples in the experiments which demonstrates the robustness of the filtering process. In contrast, a trend to produce more clusters in experiments containing few samples was observed. This was notably marked in experiments containing 10 to 15 samples. Such a bias is classical in data analysis. Indeed, if numerous values (*i.e.* samples) are used to estimate the expression profile of a given gene, outliers will have weak impact on distance calculation and the gene will be assign to the expected cluster. In contrast, when only few values are available, each of them has a greater impact on distance calculation. This results in producing more clusters with some of them having centers close to one another. This bias is also presumably amplified by the fact that small sample sets contain most generally a greater biological diversity compared to large sample sets as they contain fewer replicates. Overall, our analysis of GPL96 related experiments gave rise to 3,377 TS. The full analysis on the 33 Affymetrix platforms produced 18,250 TS which correspond to 220 millions of expression values. Partitioning results where manually checked for a large panel of experiments. Although, results seemed perfectible in few cases, they always appeared to be rational.

The TBrowser interface

Comprehensive information on samples, experiments, probes and genes were stored in a mySQL relational database. A flat file indexed on TS IDs was used to store TS expression data. This solution was preferred because it turned out to be an excellent alternative to database for retrieving rapidly expression values for the selected TS. We next developed TBrowser, a Multitier architecture system composed of (i) a "heavy client" written in JAVA (presentation Tier), (ii) a servlet container (logic tier) and (iii) a back-end database (data tier). The client application allows user

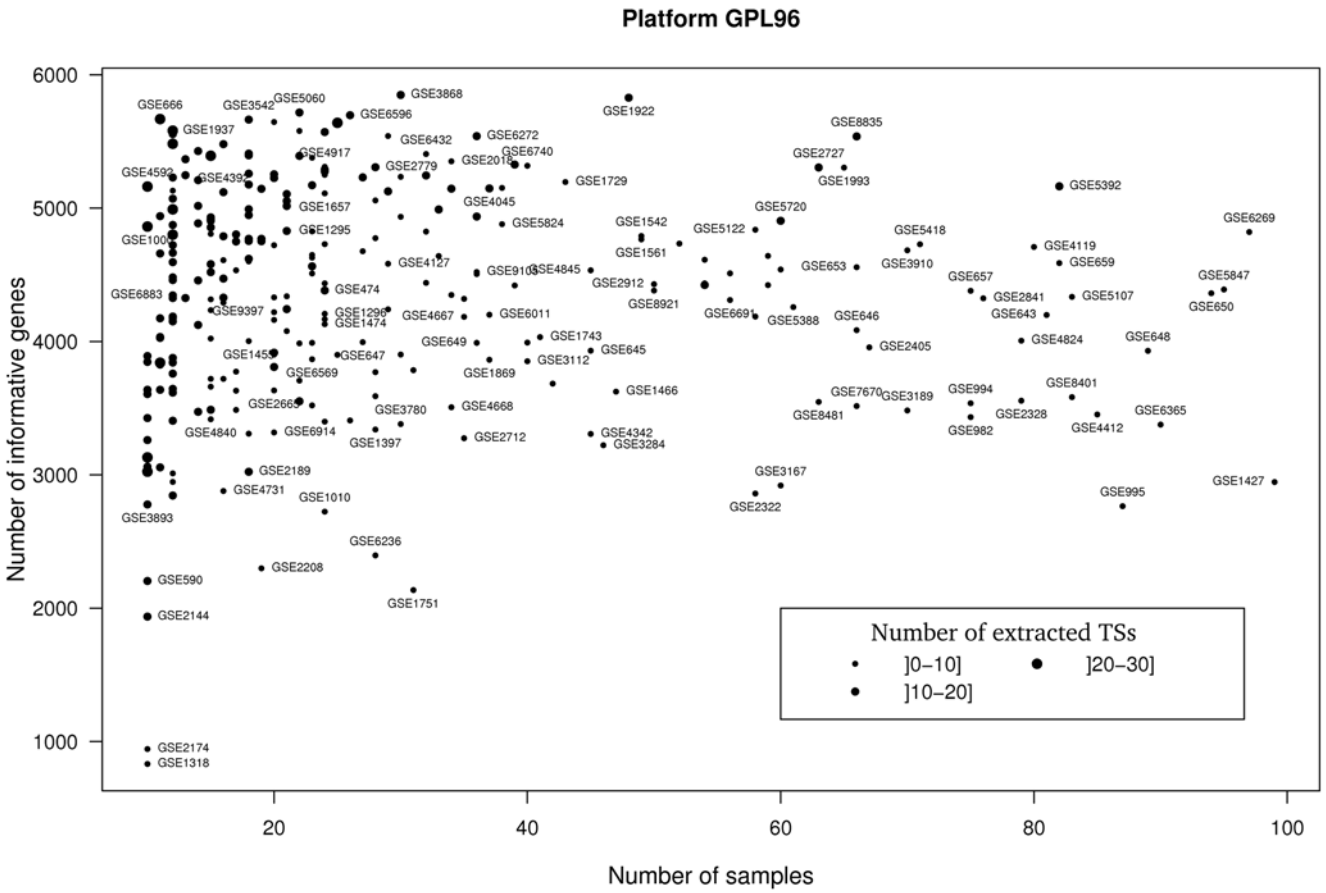


Figure 2. Large scale TS extraction from GPL96 experiments. DBF-MCL was run with default parameters (k = 100, FDR = 10%, $S_{1..3}$, Inflation = 2). X axis corresponds to the number of samples in the experiment and Y axis to the number of informative genes. For each experiment, the number of associated TS is represented by the size of the dot. For clarity purpose only experiments with less than 100 samples are represented. Furthermore, the name of only some of them is displayed.
doi:10.1371/journal.pone.0004001.g002

to query TBrowser database using six methods: by gene symbols, by probe IDs, by experiments, by microarray platform, by ontology terms (annotation) or by TS. Three of them (gene symbols, probe IDs, and annotation methods) accept a list of operators that control the way a query is to be processed. One may take advantage of these operators to create complex queries using the AND operator (&), the OR operator (|), the NOT operator (!) or using additional characters such as the quote or parenthesis (reader may refer to the user guide for additional explanations and informations). The main window of TBrowser is made of five panels (Fig. 3). The search panel is the main entry as it is used (i) to define the search method, (ii) to write the queries, (iii) to launch database interrogation and (iv) eventually to filter out some of the TS. Filters can be applied to select species of interest and to control the sizes (number of samples and number of genes) of the TS that one wants to analyze. The results area can display two panels: the list of queries the user launched during his session and the list of TS that correspond to the currently selected query. Double-clicking on one (or several) TS send it (them) to the selected plugin. The information area is used to display various informations about the selected TS whereas the plugin area is used to select one of the currently installed plugins. Finally, the plugin display panel manages the display of the currently selected plugin. To date, eight plugins have been developed (three of them are presented in this article). The Heatmap plugin is composed of two main panels: the heatmap on the left and the annotation panel on the right (Fig. 3). The Heatmap panel displays a color-coded image of TS expression values. In this representation, each row corresponds to a probe and each column to a sample. Additional informations, such as external links, can be retrieved by single-click on genes or samples. Functional enrichment informations are available on the right. The TBCommonGenes plugin was developed to compare gene composition of several TS and will be presented in the next section. Finally the TBMap plugin that can be used to visualize a summary of transcriptional regulation events observed in a given microarray platform will be presented in the last paragraph of the results section.

Meta-analysis of public microarray data using TBrowser: a case study

TBrowser can be used in many biological contexts to point out relevant experiments and construct robust gene networks. Several peer-reviewed publications have highlighted the joint regulation of the estrogen receptor-a (ESR1/ER-α), GATA3 and FOXA1 in breast cancer cells [14]. Although some of these reports have associated entry in the GEO database, retrieving neighbors of GATA3, FOXA1 and ESR1 remains a time consuming and difficult task using existing tools. As a consequence, these informations are reserved to those with strong bioinformatics skills although they are of primary interest to the biologist. Using the TBrowser search engine, this task can be translated into a very simple Boolean query, "ESR1 & GATA3 & FOXA1", which will

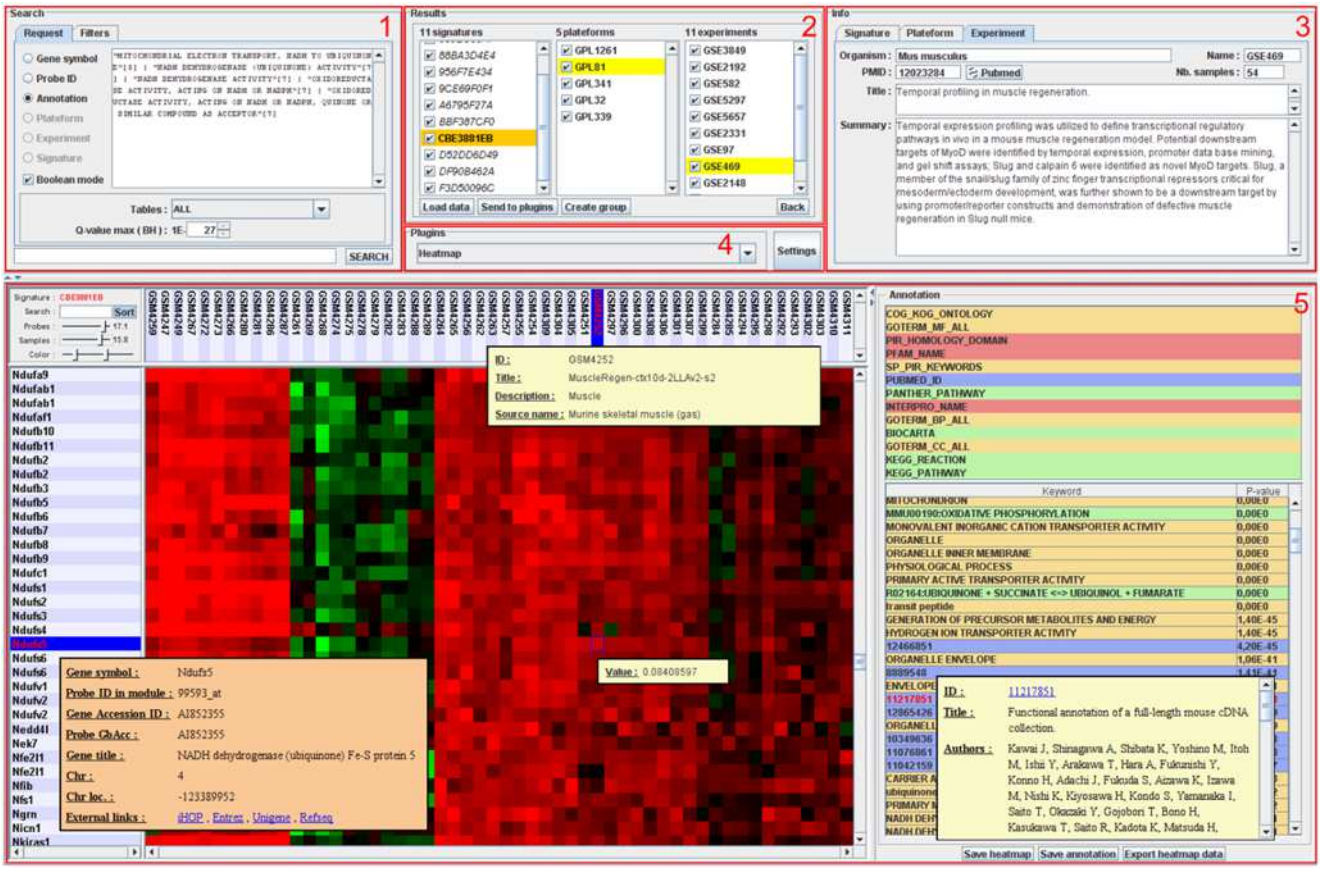


Figure 3. The TBrowser 2.0 interface. The main window of TBrowser is made of five panels (highlighted in red): the search panel (1), the results panel (2), the information panel (3), the plugins panel (4) and the plugin display panel (5). This example shows the expression profiles of genes contained in the TS CBE3881EB derived from GSE469 (''Temporal profiling in muscle regeneration''). Annotation panel shows that this TS is highly enriched in genes related to ATP synthesis.
doi:10.1371/journal.pone.0004001.g003

be almost instantaneously proceeded by the server. With the current database release, this produces a list of 16 TS (see Table 1) containing on average 508 probes (range: 82–1,572) and which were obtained using various microarray platforms (GPL96, GPL570, GPL91). Interestingly, all these TS are related to experiments performed on breast cancer cells underlying the high specificity of this gene list (Table 1). The TBCommonGenes plugin indicates that in addition to ESR1, GATA3 and FOXA1 two genes (ANXA9 and ERBB4) are found in all 16 TS. Importantly, 63 genes are found in at least 10 out of the 16 selected TS (63%). As expected, this list contains numerous markers of breast cancer cells whose expression specificity was previously reported by other (notably ERBB3, XBP1, KRT18, IL6ST, CREB1, TFF1, TFF3; see Supplementary Table S3). Thus TBrowser can be used to perform meta-analysis of microarray data in a platform-independent manner providing high confidence gene lists. However, one can also focus the analysis on a unique platform. Indeed, the transcriptional signatures 3DE64836D, B79B1C0B9 and E2E620F40 that were derived from the GPL570 platform (which measures over 47,000 transcripts) share a list of 68 genes. Many of them correspond to poorly characterized genes (for example, C17orf28 C1orf64, KIAA1370, KIAA1467, LOC143381, LOC400451, LOC92497 and ZNF703). This example clearly demonstrates the superiority of TBrowser over conventional approaches as it can be used, easily and productively, to create robust sets of transcriptionally related genes whose subsequent analysis may be crucial in defining new therapeutic targets.

Using annotation terms to mine public microarray data

Based on the systematic functional enrichment analysis, the vast majority of TS (84%) have a set of associated biological terms (only functional enrichment with q-value$<$0.01 are stored in the database). One can search for TS related to functional terms of the DAVID knowledgebase (*e.g.* ''nervous system development''). More interestingly, multiple terms can be combined with Boolean operators. Searching for TS which contain genes located in the 6p21.3 and 14q32.33 chromosomal regions (major histocompatibility complex and human immunoglobulin heavy-chain locus respectively) and which contain T-cell specific genes, can be translated as: 6p21.3[4] & 14q32.33[4] & ''T CELL ACTIVATION''[5,12] ([4] = cytoband term, [5] = GO term, [12] = Panther pathways term). As chromosomal aberrations do occur frequently in cancer our approach can also be used to perform systematic cytogenetic analysis. Indeed, throughout our analysis, 2,208 functional enrichments related to 360 human cytobands were observed and stored in the database. As an example, TS with very strong enrichment (q-value$<1.10^{-20}$) for any of the human cytobands stored in the database are presented in Table 2. The first one is related to atopic dermatis analysis (skin biopsies) and contained 24% of genes located in 17q12-q21. They correspond to genes encoding for the keratin and keratin-associated protein families (KRT17, KRT27, KRTAP1-5, KRTAP17-1, KRTAP3-1, KRTAP3-3, KRTAP4-10, KRTAP4-12, KRTAP4-13, KRTAP4-15, KRTAP4-2, KRTAP4-3, KRTAP4-5, KRTAP4-8, KRTAP4-9, KRTAP9-2, KRTAP9-3, KRTAP9-4 and

Table 1. Transcriptional signatures containing Affymetrix probes for ESR1, GATA3 and FOXA1.

TS ID[1]	Genes[2]	Probes[2]	Samples[2]	Sample type	GSE ID	GPL ID	Author	PubMed IDs
0F2635383	1190	1572	23	Cell lines	GSE6569	GPL96	Huang F et al 2007	17332353
3DE64836D	102	143	62	Tissue	GSE7904	GPL570	unpublished 2007	-
59A18E225	690	893	121	Both	GSE2603	GPL96	Minn AJ et al 2005	16049480
6C975B20B	88	96	26	Tissue	GSE6772	GPL96	Klein A et al 2007	17410534
6C975B290	88	96	26	Tissue	GSE6596	GPL96	Klein A et al 2007	17410534
7150E17F6	868	1032	34	Cell lines	GSE4668	GPL96	Coser KR et al 2003	14610279
8059848B4	200	250	251	Tissue	GSE3494	GPL96	Miller LD et al 2005	16141321
84E5E1077	694	883	198	Tissue	GSE7390	GPL96	Desmedt C et al 2007	17545524
8F69864F9	68	82	95	Tissue	GSE5847	GPL96	Boersma BJ et al 2007	17999412
A151D5695	297	361	58	Tissue	GSE5327	GPL96	Minn AJ et al 2007	17420468
B79B1C0B9	270	380	47	Tissue	GSE3744	GPL570	Richardson AL et al 2006	16473279
BDB6D8700	550	679	104	Tissue	GSE3726	GPL96	Chowdary D et al 2006	16436632
D8F0B528C	125	152	159	Tissue	GSE1456	GPL96	Pawitan Y et al 2005	16280042
E2E620F40	448	616	129	Tissue	GSE5460	GPL570	unpublished 2007	-
EA9669A21	219	251	158	Tissue	GSE3143	GPL91	Bild AH et al 2006	16273092
F310ACC36	519	646	49	Tissue	GSE1561	GPL96	Farmer P et al 2005	15897907

[1]Transcriptional signature ID.
[2]Total number.
doi:10.1371/journal.pone.0004001.t001

KRTAP9-8). This signature is notably annotated as being enriched in genes related to PMID 11279113 (“Characterization of a cluster of human high/ultrahigh sulfur keratin-associated protein genes embedded in the type I keratin gene domain on chromosome 17q12-21”) [15] and in genes related to the PIR keyword “multigene family”. Furthermore, several signatures, of Table 2 are related to melanoma and six of them were observed in the GSE7127 experiment [16]. Although data from Table 2 would deserve further analysis they are most likely related to gain or loss of genetic material in tumors. Indeed, gain of 8q is frequently observed in a number of tumor types (including melanoma and ovarian tumors) and this region is known to contain the c-myc oncogene at 8q24.21. Interestingly, in several cases, contiguous cytobands were significantly enriched suggesting a large deletion or amplification of genetic material in these tumors (TS 60E29DA83 is enriched in genes from 8q13, 8q21.11, 8q22.1, 8q22.3, 8q24.13 and 8q24.3 cytobands). In the same way, loss of genetic material of the long arm of chromosome 11 occurs in primary melanoma but is even more frequent in metastatic tumors (TS A93ED7519 is enriched in genes from 11q21, 11q23.3 and 11q24.2 cytobands). Altogether, these results underline the versatility of TBrowser and its ability to extract hidden and meaningful informations from published or unpublished micro-array data. Indeed, the cytogenetic results presented in Table 2 were not discussed by the authors in the corresponding articles.

A synthetic view of all GPL96 related experiments

The paradigm that genes from a TS share functional relationships is now widely accepted and constitutes the basis of transcriptome analysis [17]. However, each of these TS is rather

Table 2. Transcriptionnal signatures displaying high enrichment (q value$<1.10-20$) for any of the human cytoband tested.

TS ID[1]	Enrich.[2]	Cytoband	q.value	Sample type	GSE ID	GPL ID	Authors	PubMed ID
3DA3C8345	24%	17q12-q21	$1.7.10^{-39}$	Skin	GSE5667	GPL97	Plager DA et al 2007	17181634
43CC3EF57	9%	8q24.3	$7.0.10^{-32}$	Melanoma	GSE7153	GPL570	Unpublished 2007	-
60E29DA83	16%	8q24.3	$6.8.10^{-24}$	Melanoma	GSE7127	GPL570	Johansson P et al 2007	17516929
60E581184	26%	17q25.1	$5.5.10^{-23}$	Melanoma	GSE7127	GPL570	Johansson P et al 2007	17516929
60E6B4129	35%	20p13	$1.6.10^{-26}$	Melanoma	GSE7127	GPL570	Johansson P et al 2007	17516929
60E96FF1E	28%	6p21.3	$1.2.10^{-28}$	Melanoma	GSE7127	GPL570	Johansson P et al 2007	17516929
60EC95F6A	17%	7q22.1	$6.3.10^{-31}$	Melanoma	GSE7127	GPL570	Johansson P et al 2007	17516929
60EEBD669	32%	11q23.3	$1.4.10^{-26}$	Melanoma	GSE7127	GPL570	Johansson P et al 2007	17516929
B4C95CF18	42%	8q24.3	$1.1.10^{-36}$	Ovary	GSE6008	GPL96	Hendrix ND et al 2006	16452189
A93ED6519	16%	11q23.3	$6.9.10^{-23}$	Melanoma	GSE7152	GPL570	Packer LM et al 2007	17450523
A93DB01ED	11%	7q22.1	$9.5.10^{-30}$	Melanoma	GSE7152	GPL570	Packer LM et al 2007	17450523

[1]Transcriptional signature ID.
[2]Enrichment: Proportion of non redondant genes from the TS that are located in the corresponding cytoband.
doi:10.1371/journal.pone.0004001.t002

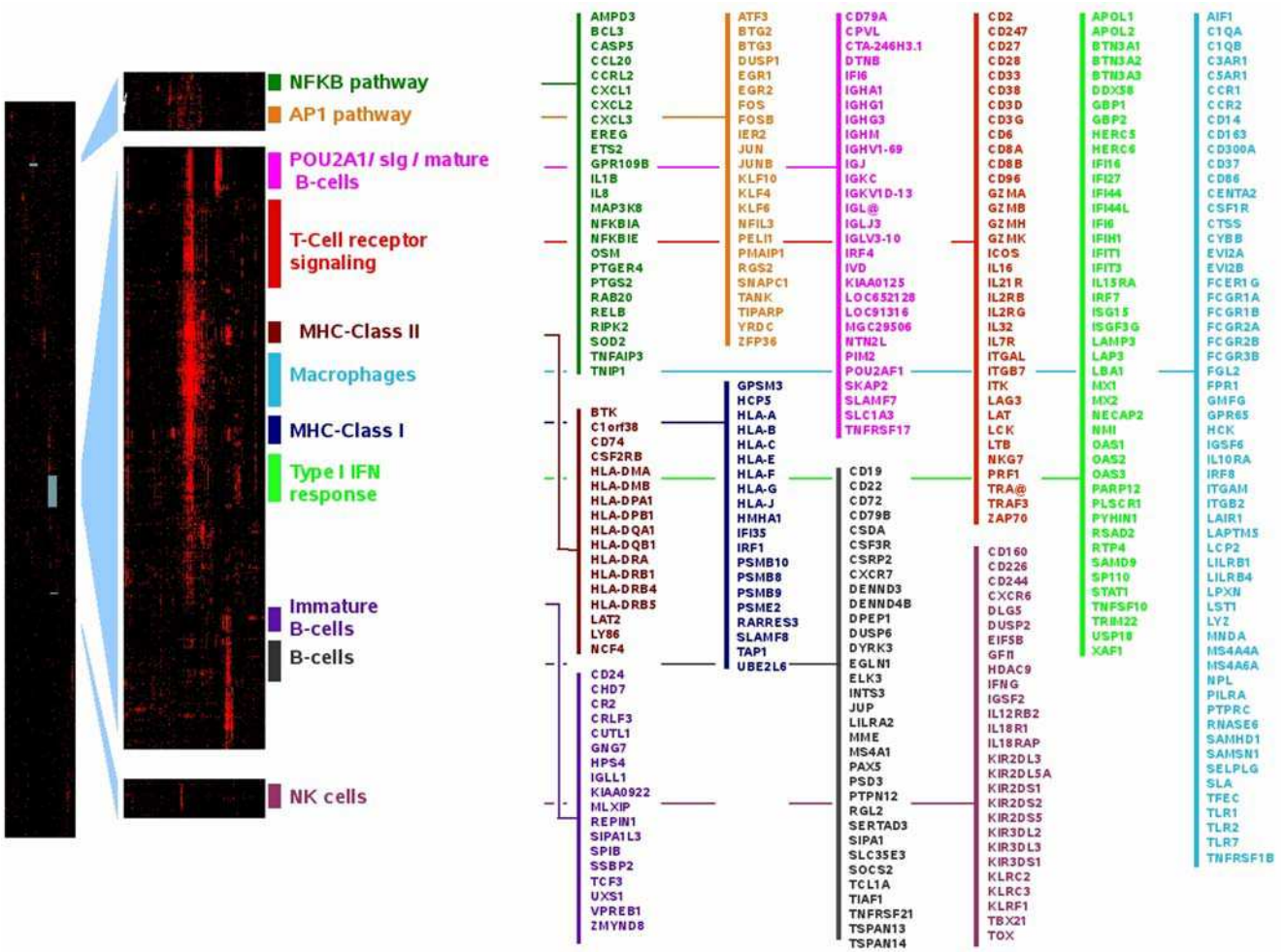


Figure 4. The transcriptional MAP associated with GPL96 related experiments. (A) A low resolution image made of 22,215 probes from GPL96 platform as rows and 3,114 GPL96 specific TS as columns. Red color indicates the presence of a gene in the corresponding TS (default to black). **(B)** Zooms of the corresponding areas showing some immune system related meta-signatures. **(C)** Representative genes that fall into these clusters.
doi:10.1371/journal.pone.0004001.g004

associated to multiple underlying pathways whose components and limits are unclear. Our difficulty in depicting comprehensive maps for pathways is illustrated by existing discrepancies, for instance, between those proposed by BioCarta, KEGG and GeneMAPP. We reasoned that the more frequently two genes fall in the same TS, the more likely these genes belong to the same core functional network. To test this hypothesis, we produced a Boolean matrix with 22,215 probes from GPL96 platform as rows and 3,114 GPL96 specific TS as columns (only TS containing 30 to 1500 probes were included). This matrix was filled with zero and elements were set to 1 if a given gene was observed in the corresponding TS. Hierarchical clustering with uncentered Pearson's correlation coefficient was used to reveal genes frequently associated to the same TS. Given the order of the resulting matrix, it could not be visualized on a desktop computer using conventional software (*i.e.*; Treeview, MeV). We thus developed the TBMap plugin which allows one to visualize the map but also to superimpose a user-defined or a KEGG-related gene list. As expected, most of the clusters where obviously enriched in genes involved in similar biological processes (Protein biosynthesis/Ribosome function, oxidative phosphorylation, cell cycle, fatty acid metabolism, valine leucine and isoleucine degradation, extracellular matrix, breast cancer cells, structural constituent of muscles, neuronal processes, etc.). This was particularly clear when KEGG pathway informations were superimposed (see Figure S6). The Figure 4 presents some of the clusters that were identified as related to immune system functions. We could find a signature defining T cells that contained numerous cell-surface markers (e.g. TCA@, CD2, CD3G, CD6, IL2RB, IL2RG, IL7R, IL21R and ICOS), signaling genes (ZAP70, LAT, LCK, ITK) and cytotoxicity-related genes (GZMA, GZMB, GZMH, GZMK and PRF1). Concerning B-cells, three clusters were observed. A large signature contains mature B-cell markers (CD19, CD22, CD72 and CD79B) and transcription factors important in B-cell development such as PAX5 and TCL1A. A second signature contains POU2AF1/OBF-1, together with its described targets: genes coding for immunoglobulin (IGHG1, IGHG3, IGHA1, IGHM, IGJ, IGKC and IGL) and the B-cell maturation factor, TNFRSF17/BCMA [18,19]. The third B-cell signature contains cell surface markers found in immature B-cells (CD24, VPREB1, IGLL1/CD179B and CR2/CD21) in addition to transcription factors known to play a crucial role during early B-cell development (TCF3, SPIB and CUTL1). The NK signature contains eight genes of the Killer cell immunoglobulin-like receptors (KIR) family, 3 genes of the killer cell lectin-like receptor family in addition to other markers whose expression has

been reported on the surface of NK cells (CD160, CD244/2B4 and CD226) [20,21,22]. It also contains TBX21/T-bet together with IL18R1, IL18RAP, IL12RB2 and IFNG. Importantly, the IL12/IL18 combination has been shown to be potent inducers of both TBX21/T-bet and IFNG in NK cells[23,24]. In addition to MHC-Class I, MHC-Class II and macrophage related signatures, two pathways related to immune function are presented in Figure 4. The AP1 pathway is made of the prototypical immediate early genes and contains numerous transcription factors (EGR1, EGR2, FOS, FOSB, IER2, JUN, JUNB, KLF6, KLF4, KLF10, ATF3, BTG2 and BTG3) whose complex interplay has been reported earlier. Finally, a NFKB signature was also observed which, again, contains prototypical regulators (NFKIA, NFKIE, RELB, BCL-3 and MAP3K8/TPL2) and known targets (CCL20, CXCL3, IL1B, IL8 and SOD2). Altogether, these results underline the high relevance of the signatures obtained using this compilation of TS derived from GPL96 related GEO experiments.

Discussion

In the present paper, we present the construction of a unique collection of TS that summarize almost all human, mouse and rat Affymetrix microarray data stored in the GEO database. TBrowser constitutes a highly powerful search engine that makes it possible to perform easily platform independent meta-analysis of microarray data. This can be considered as a real improvement over classical approaches and softwares as it provides easy and productive access to data without the need of any programming skills. Indeed the simple use of an extended set of operators proved to be sufficient to construct robust gene networks and assign poorly characterized genes to relevant biological pathways. As a consequence, it is particularly well suited to compare results obtained through microarray, ChIP-on-chip, ChIP-seq, CGH or protein-protein interaction experiments to those previously stored in the GEO database.

In all tested experiments, we found that DBF-MCL gives very good results both on simulated datasets and real microarray datasets. Although Lattimore *et al* proposed another MCL-based algorithm (geneMCL) we were unable to compare our results with their implementation as the software is no longer available nor maintained. However, DBF-MCL was run on the full van't Veer DataSet [25] (117 biological samples) that was used by Lattimore and collaborators in the original paper. In their report, the authors used a subset of genes (5,730 out of 24,482) that were selected based on their associated variance. Our procedure run on the full dataset led to the selection of 5,932 genes that fall into 22 clusters (in contrast to 154 clusters using geneMCL). This discrepancy is likely to be due to the filtering step applied to the dataset. Indeed, a strong associated variance can also be reminiscent of punctual random artifacts. Thus, selecting those genes will generate small or singleton clusters. In this context, the MDNN statistic better handle these artifacts as its purpose is to conserve genes that belong to dense region in the hyperspace.

To date, TBrowser provides user with only one partitioning solution for a dataset. However as density is heterogeneous inside a dataset, several partitioning solutions exist. For instance, if one observes a cluster containing cells of the immune system this will also frequently contain several sub-clusters that will be reminiscent of cell types (B- or T-cells for example) or activation status. Increasing MCL granularity ("Inflation" parameter) will most generally split the parent clusters and provided user with another partitioning result. However, both results can be considered as optimal and we should consider all of them. To this end we plan to propose multiple partitioning solutions for each dataset to provide a more exhaustive view of underlying biological pathways. Although, such an approach could appear computer-intensive it should be practicable, taking into account that DBF-MCL is much faster than hierarchical clustering or MCL run on a whole dataset. In addition, although we routinely obtained very relevant results with DBF-MCL, we expect that even more accurate methods will be proposed in the future.

The present work focus on human, mouse and rat Affymetrix microarray data but TBrowser can handle any type of microarrays and organism. The current release of the database already contains data obtained using other commercial (*e.g.* Agilent, Illumina Inc., GE Healthcare, Applied Biosystems, Panomics, CapitalBio Corporation, TeleChem ArrayIt, Mergen-LTD, Eppendorf Array Technologies) and non commercial platforms (*e.g.* National Cancer Institute, Vanderbilt Microarray Shared Resource, Genome Institute of Singapore), several of them being related to the MicroArray Quality Control (MAQC) project (GSE5350) [26]. However, to date, systematic analysis of all experiments performed on these platforms has not been done. The flexibility of our approach also makes it possible to integrate and compare data obtained through any kind of large scale analysis technologies providing that the experiment can be represented by a single numerical matrix (ChIP-on-chip, Protein array, large scale Real-time PCR, ChIP-seq, etc.). Three plugins (Heatmap, TBCommonGenes and TBMap) have been presented in this article but seven new plugins have been recently developed (manuscript in preparation). In the near future, the ease of plugin development will makes it possible to look for TS enriched in genes sharing transcription factor and miRNA specific motifs in their non-coding regions.

As raw data are only available for some of the microarray datasets, we used the "normalized" data provided by submitters. These data were subsequently rank-transformed and used for classification. This procedure allowed us to re-analyze a very large number of datasets. However, the drawback is that quality status of individual samples or experiments could not be determined (computing the so-called "3′/5′ ratio" requires raw data). We plan to provide extensive quality control informations through a dedicated plugin. However, we think that scientists should comply better with the MIAME guidelines and that they should provide systematically raw data when submitting a new experiment. Finally, we would like to acknowledge the GEO database team whose efforts in providing high quality repository service made this work possible.

Materials and Methods

Microarray data retrieval

Human mouse and rat microarray data derived from 30 Affymetrix microarray platforms (Supplementary Table S1) were downloaded from the GEO ftp site and retrieved in seriesMatrix file format (ftp://ftp.ncbi.nih.gov/pub/geo/DATA/SeriesMatrix/). SeriesMatrix are summary text files related to a GEO series Experiment (GSE) that include sample and experiment metadata together with a tab-delimited matrix that corresponds to normalized expression data. Each file (n = 2,869) was parsed using a Perl script to extract gene expression matrix and metadata. Probes with missing expression values were excluded from analysis. Only expression matrix with at least ten columns/samples were kept for subsequent analysis (n = 1,484, Supplementary Table S2).

DBF-MCL algorithm

The filtering step of DBF-MCL was implemented in C. The latest Markov Clustering algorithm version (1.006, 06-058) was

obtained from http://micans.org/mcl/src/. The full pipeline of DBF-MCL (that integrates normalization, filtering and partitioning) was implemented in Bash Shell Scripting language. This script supports different metrics for distance calculation (Euclidean distance, Pearson's correlation coefficient-based distance, Spearman's rank correlation-based distance).

Data normalization and processing

Given the huge amount of data processed by GEO curators it is impractical to determine the quality and efficiency of the normalization methods used [27]. Although seriesMatrix files should ideally contain log-transformed data, expression matrices in linear scale were also observed in several cases. To circumvent this problem each column of the expression matrix was rank-transform (using R software). This normalization procedure is insensitive to data distribution and provided us with a standard input for the DBF-MCL algorithm. In the case of microarray data, DBF-MCL was run using Spearman's rank correlation-based distance (1-r). However, although rank-based methods are well suited for normalization and distance calculation purposes they are not appropriate to display gene expression profiles. To this end, a normal score transformation was applied to each column of the datasets after DBF-MCL classification. The transformation ensures that whatever the data a standard format is available for heatmap visualization. Finally, for each experiment, this dataset was used (1) to classify samples using hierarchical clustering (2) to build the expression matrix for the corresponding TS.

Data storage

Expression matrix for each TS were stored in an indexed flat file with a TS ID as a key. This flat file is used by the TBrowser client to retrieve expression data for the requested TS. Experiment metadata, corresponding to sample and experiment informations were stored in a mySQL relational database. Probe meta-information (gene symbol, gene name, GenBank accession ID, chromosomal location, Entrez ID) were obtained from Bioconductor [28] annotation packages and stored in the database. In some cases, as no annotation packages were available (especially for GeneChip® CustomExpress® Array) a script was used to obtain gene symbols and gene names from GenBank files based on the provided GenBank accession ID. Both flat file and database information will be periodically updated to give access to novel experiments stored in GEO repository.

Complex9 dataset

The complexe9 dataset was obtained from the UH Data Mining and Machine Learning Group (UH-DMML, http://www2.cs.uh.edu/~ml_kdd/). Cluster Affinity Search Technique (CAST) was run using the TMEV software. QT_CLUST and k-means were run using the flexclust and fpc R package. For k-means, the algorithm was run 10 times with random initial centers. Hierarchical clustering was performed using the amap library from the R/Bioconductor project. The Euclidean distance was used in all cases.

Functional enrichment analysis

We used the DAVID knowledgebase [12] for functional enrichment analysis as it provided a practical mean to gain access to a wide range of heterogeneous sources of gene annotation (152,543 annotation terms were used for human, 105,207 for mouse and 39,787 for rat). DAVID ID mapping was obtained for 218,727 AffyID. A Perl script that integrates call to the R software was run to load probe list and calculate iteratively Fisher's exact test p-values on 2×2 contingency tables. Bonferroni adjusted p-values were calculated using the multtest Bioconductor library for all TS. Overall, 5.10^6 Fisher's exact test were performed.

User interface

TBrowser is accessible through a web browser at TAGC web site (http://tagc.univ-mrs.fr/tbrowser/). Of note, the TBrowser client is extensible through a plug-in architecture that allows rapid development of additional features. A developer's guide will be available soon on our website.

Supporting Information

Figure S1 A schematic overview of the pipeline used in TBrowser.
Found at: doi:10.1371/journal.pone.0004001.s001 (10.16 MB TIF)

Figure S2 An illustration in two dimensions of the motivation behind DBF-MCL filtering step. Arrows point out the 20th nearest neighbor for selected points. Length of each segment corresponds to a given DKNN value.
Found at: doi:10.1371/journal.pone.0004001.s002 (8.22 MB TIF)

Figure S3 Distributions of DKNN values. Observed DKNN values (solid line) and of a set of simulated DKNN values S (dotted line) are shown for (A) the Complex9RN200 artificial dataset and (B) the GSE1456 microarray dataset.
Found at: doi:10.1371/journal.pone.0004001.s003 (9.01 MB TIF)

Figure S4 Colors correspond to the clusters found using the corresponding algorithm (A) The whole dataset (9,112 points). (B) A zoom-in of Complex9RN200 dataset that displays the various shapes to be found. (C) DBF filtering step without partitioning. With k set to 60, noisy elements remain around the shapes. (D–G) The filtering and partitioning results obtained using DBF-MCL run with a range of k values and I values. Other arguments are unchanged (FDR = 10%, S1..3). The set of points (n = 3,108) obtained using DBF-MCL (k = 20) was used to test the other algorithms (H) Results obtained with hierarchical clustering (single linkage). The obtained dendrogram was cut to produce 9 clusters. (I) Results obtained with the QT_CLUST algorithm (radius = 0.8). (J) Results obtained for k-means (9 centers, 100 initializations). (K) Results obtained with cst(threshold = 0.81).
Found at: doi:10.1371/journal.pone.0004001.s004 (9.41 MB TIF)

Figure S5 Impact of various k values on DBF-MCL results. The x-axis correspond to k values. The y-axis correspond to the number of elements considered as informative. (A) DBF-MCL was run with the Complex9RN200 as input using a range of k values (FDR = 10%, S1..3, Inflation = 1.2). (B) DBF-MCL was run with several microarray datasets as input (including GSE1456) using a range of k values (FDR = 10%, S1..3, Inflation = 2).
Found at: doi:10.1371/journal.pone.0004001.s005 (8.72 MB TIF)

Figure S6 The TBMap plugin. These pictures are derived from the GPL96 map (22,215 probes as rows and 3,114 GPL96 specific TS as columns). Red indicates the presence of a gene in the corresponding TS (default to black). Only small parts of the map are displayed. (A) A cluster enriched in genes from the "Aminoacyl-tRNA biosynthesis" KEGG pathway (hsa00970). Genes (rows) from this KEGG pathway are displayed as blue lines (CARS, SARS, AARS, GARS, MARS, IARS, YARS). Genes from a manually entered gene list are shown in yellow (TRIB3, MOCOS, MPZL1, CBS, PPCDC). (B) A cluster enriched in genes related to oxydative phosphorylation (KEGG pathway hsa00190, "Oxidative phosphorylation"). (C) A cluster containing

genes related to ribosome biogenesis (KEGG pathway hsa03010 "Ribosome"). (D) A cluster enriched in genes involved in cell proliferation (KEGG pathway hsa04110 "Cell cycle").
Found at: doi:10.1371/journal.pone.0004001.s006 (9.66 MB TIF)

Table S1 Informations related to Affymetrix platforms (n = 33) used in the present work.
Found at: doi:10.1371/journal.pone.0004001.s007 (0.12 MB XLS)

Table S2 Informations related to experiments (n = 1,484) that were analyzed using the DBF-MCL algorithm. All Informations were obtained from the GEO website.
Found at: doi:10.1371/journal.pone.0004001.s008 (1.38 MB XLS)

Table S3 This matrix summarizes the results obtained using the "ESR1 & GATA3 & FOXA1" query. Rows correspond to genes and columns to TS. The presence of a given gene in a given TS is indicated by 1 (default 0).
Found at: doi:10.1371/journal.pone.0004001.s009 (0.66 MB XLS)

Acknowledgments

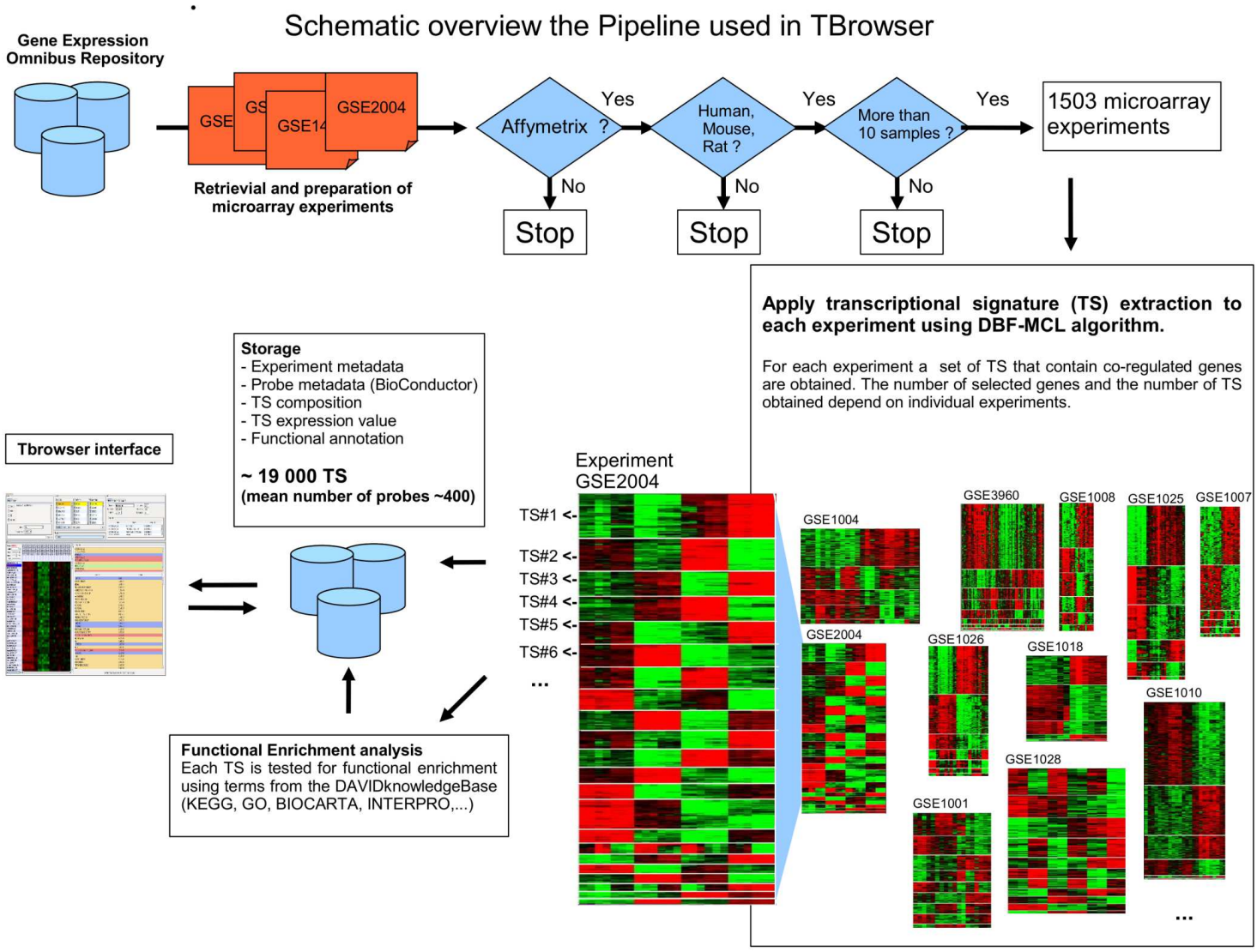
The authors would like to thanks the staff from the TAGC laboratory for helpful discussions and gratefully acknowledge Francois-Xavier Theodule for technical assistance.

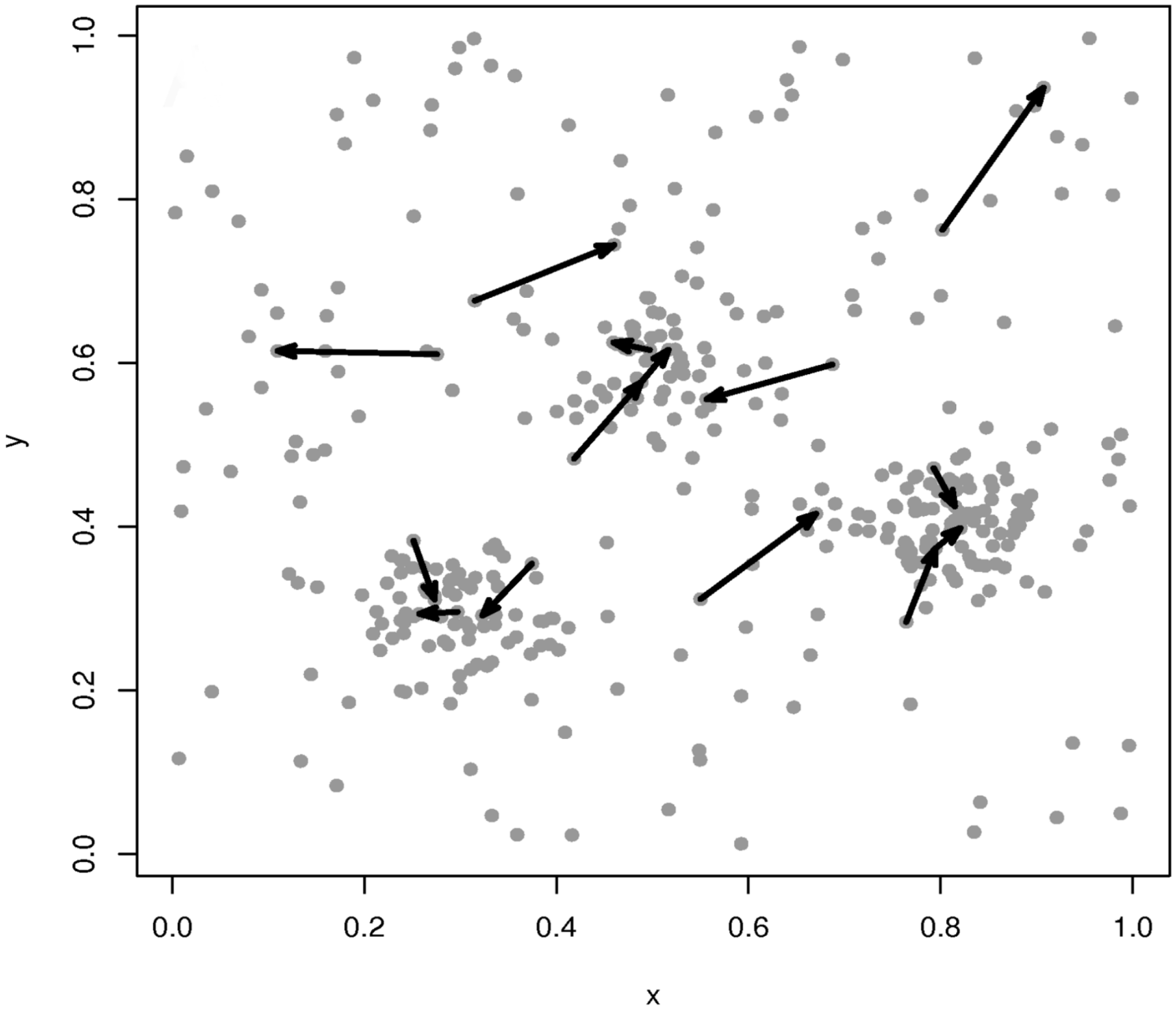
Author Contributions

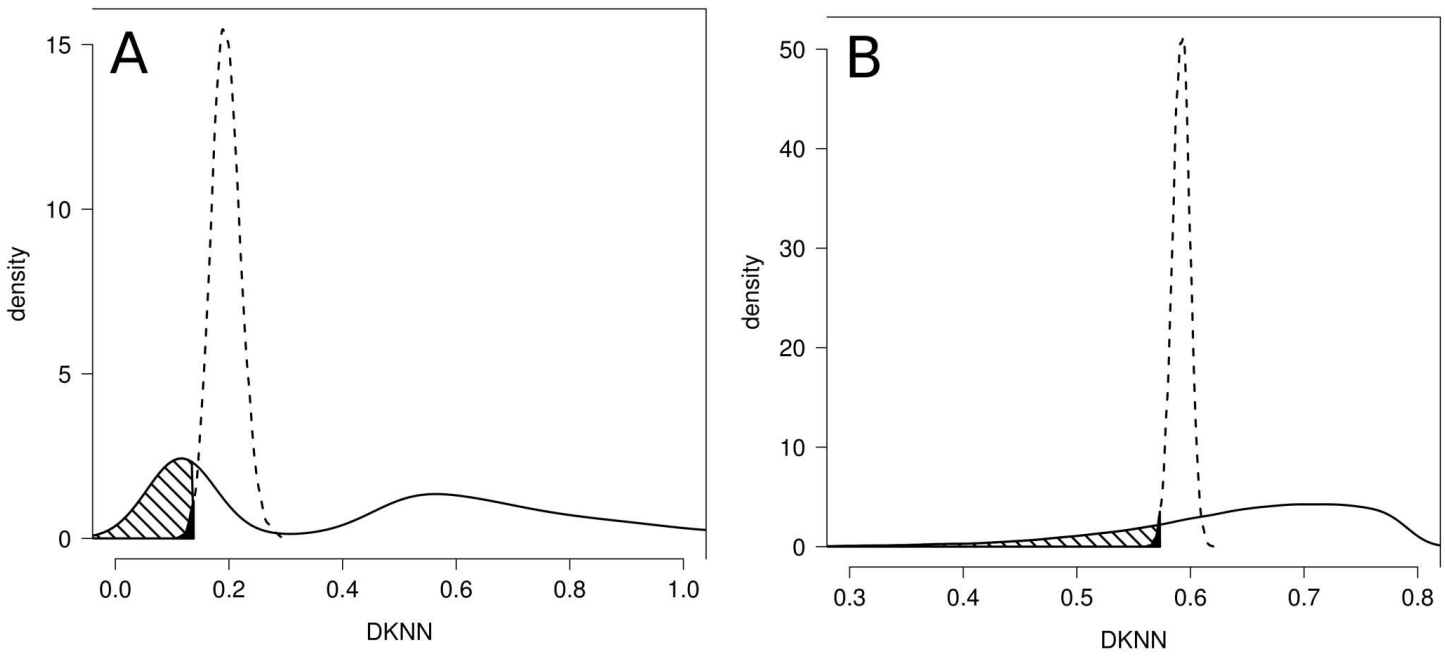
Conceived and designed the experiments: FL JT AB GD ER SG DP. Performed the experiments: FL JT AB DP. Analyzed the data: FL JT AB JI CN DP. Contributed reagents/materials/analysis tools: FL JT AB GD ER SG JI DP. Wrote the paper: JI CN DP.

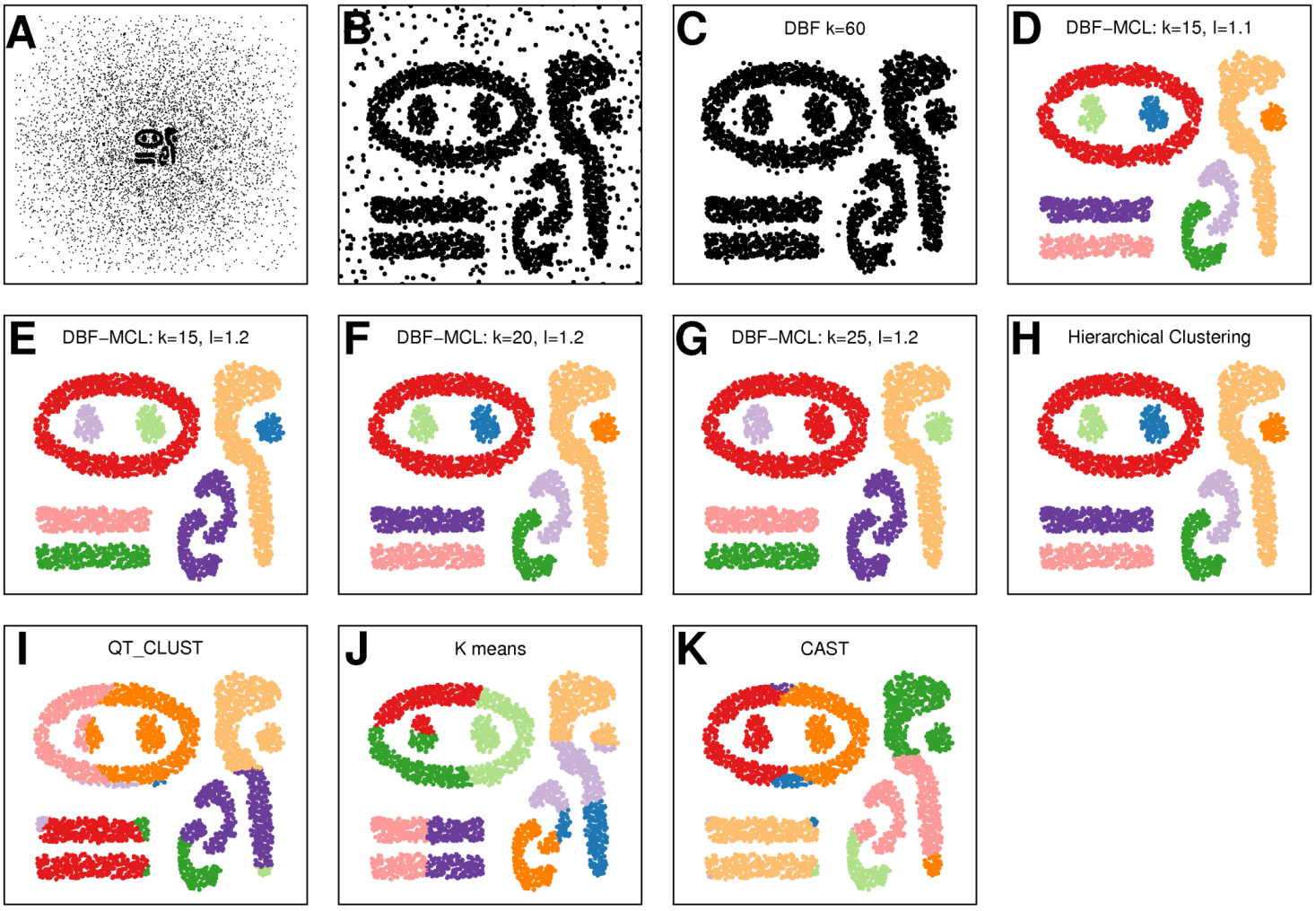
References

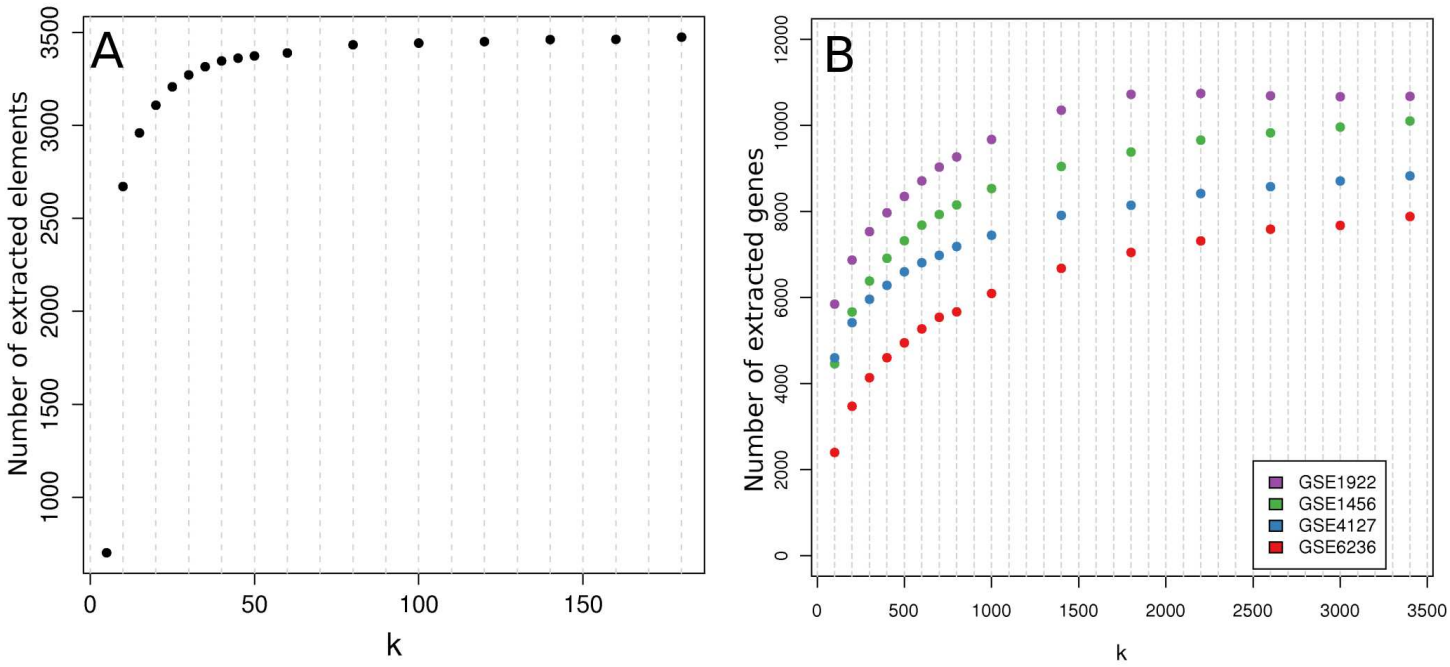
1. Stoeckert CJ, Causton HC, Ball CA (2002) Microarray databases: standards and ontologies. Nat Genet 32 Suppl: 469–73.
2. Barrett T, Edgar R (2006) Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods Enzymol 411: 352–69.
3. Diehn M, Sherlock G, Binkley G, Jin H, Matese JC, et al. (2003) SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. Nucleic Acids Res 31: 219–23.
4. Parkinson H, Sarkans U, Shojatalab M, Abeygunawardena N, Contrino S, et al. (2005) ArrayExpress–a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 33: D553–5.
5. D'haeseleer P (2005) How does gene expression clustering work? Nat Biotechnol 23: 1499–501.
6. Heyer LJ, Kruglyak S, Yooseph S (1999) Exploring expression data: identification and analysis of coexpressed genes. Genome Res 9: 1106–15.
7. CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling (n.d.) Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi = 10.1.1.44.5847. Accessed 18 September 2008.
8. Van Dongen S (2000) A cluster algorithm for graphs. National Research Institute for Mathematics and Computer Science in the. pp 1386–3681.
9. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, et al. (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 440: 637–43.
10. Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30: 1575–84.
11. Samuel Lattimore B, van Dongen S, Crabbe MJC (2005) GeneMCL in microarray analysis. Comput Biol Chem 29: 354–9.
12. Sherman BT, Huang DW, Tan Q, Guo Y, Bour S, et al. (2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics 8: 426.
13. Pawitan Y, Bjöhle J, Amler L, Borg A, Egyhazi S, et al. (2005) Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. Breast Cancer Res 7: R953–64.
14. Lacroix M, Leclercq G (2004) About GATA3, HNF3A, and XBP1, three genes co-expressed with the oestrogen receptor-alpha gene (ESR1) in breast cancer. Mol Cell Endocrinol 219: 1–7.
15. Rogers MA, Langbein L, Winter H, Ehmann C, Praetzel S, et al. (2001) Characterization of a cluster of human high/ultrahigh sulfur keratin-associated protein genes embedded in the type I keratin gene domain on chromosome 17q12-21. J Biol Chem 276: 19440–51.
16. Johansson P, Pavey S, Hayward N (2007) Confirmation of a BRAF mutation-associated gene expression signature in melanoma. Pigment Cell Res 20: 216–21.
17. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863–8.
18. Strubin M, Newell JW, Matthias P (1995) OBF-1, a novel B cell-specific coactivator that stimulates immunoglobulin promoter activity through association with octamer-binding proteins. Cell 80: 497–506.
19. Zhao C, Inoue J, Imoto I, Otsuki T, Iida S, et al. (2008) POU2AF1, an amplification target at 11q23, promotes growth of multiple myeloma cells by directly regulating expression of a B-cell maturation factor, TNFRSF17. Oncogene 27: 63–75.
20. Rabot M, El Costa H, Polgar B, Marie-Cardine A, Aguerre-Girr M, et al. (2007) CD160-activating NK cell effector functions depend on the phosphatidylinositol 3-kinase recruitment. Int Immunol 19: 401–9.
21. Boles KS, Nakajima H, Colonna M, Chuang SS, Stepp SE, et al. (1999) Molecular characterization of a novel human natural killer cell receptor homologous to mouse 2B4. Tissue Antigens 54: 27–34.
22. PubMed Snapshot (n.d.) Available: http://www.ncbi.nlm.nih.gov/sites/entrez. Accessed 18 September 2008.
23. Townsend MJ, Weinmann AS, Matsuda JL, Salomon R, Farnham PJ, et al. (2004) T-bet regulates the terminal maturation and homeostasis of NK and Valpha14i NKT cells. Immunity 20: 477–94.
24. Lauwerys BR, Renauld JC, Houssiau FA (1999) Synergistic proliferation and activation of natural killer cells by interleukin 12 and interleukin 18. Cytokine 11: 822–30.
25. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415: 530–6.
26. Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, et al. (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nat Biotechnol 24: 1151–61.
27. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, et al. (2007) NCBI GEO: mining tens of millions of expression profiles–database and tools update. Nucleic Acids Res 35: D760–5.
28. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.


Schematic overview the Pipeline used in TBrowser
Gene Expression Omnibus Repository
GSE
GS
GSE14
GSE2004
Retrievial and preparation of microarray experiments
Affymetrix ?
Yes
No
Stop
Human, Mouse, Rat ?
Yes
No
Stop
More than 10 samples ?
Yes
No
Stop
1503 microarray experiments
Apply transcriptional signature (TS) extraction to each experiment using DBF-MCL algorithm.
For each experiment a set of TS that contain co-regulated genes are obtained. The number of selected genes and the number of TS obtained depend on individual experiments.
Storage
- Experiment metadata
- Probe metadata (BioConductor)
- TS composition
- TS expression value
- Functional annotation
~ 19 000 TS
(mean number of probes ~400)
Tbrowser interface
Experiment GSE2004
TS#1 <-
TS#2 <-
TS#3 <-
TS#4 <-
TS#5 <-
TS#6 <-
...
GSE1004
GSE3960
GSE1008
GSE1025
GSE1007
GSE2004
GSE1026
GSE1018
GSE1010
GSE1028
GSE1001
...
Functional Enrichment analysis
Each TS is tested for functional enrichment using terms from the DAVIDknowledgeBase (KEGG, GO, BIOCARTA, INTERPRO,...)

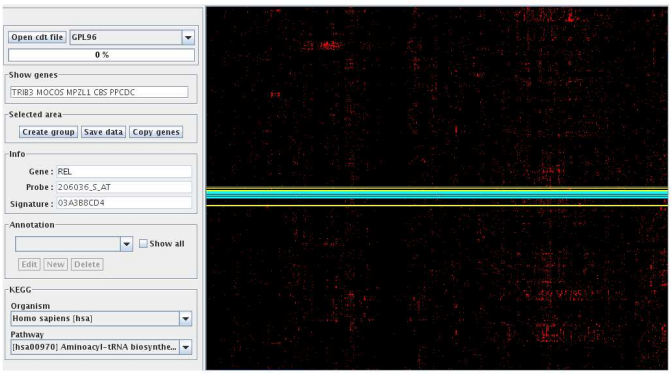

y
x
0.0
0.2
0.4
0.6
0.8
1.0


A
B
density
DKNN
0
5
10
15
20
30
40
50
0.0
0.2
0.4
0.6
0.8
1.0
0.3
0.4
0.5
0.6
0.7
0.8

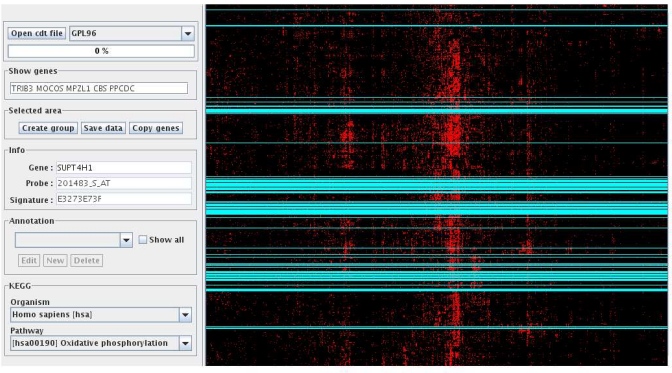

A
B
C
DBF k=60
D
DBF−MCL: k=15, I=1.1
E
DBF−MCL: k=15, I=1.2
F
DBF−MCL: k=20, I=1.2
G
DBF−MCL: k=25, I=1.2
H
Hierarchical Clustering
I
QT_CLUST
J
K means
K
CAST


A
Number of extracted elements
1000
1500
2000
2500
3000
3500
0
50
100
150
k
B
Number of extracted genes
0
2000
4000
6000
8000
10000
12000
0
500
1000
1500
2000
2500
3000
3500
k
GSE1922
GSE1456
GSE4127
GSE6236

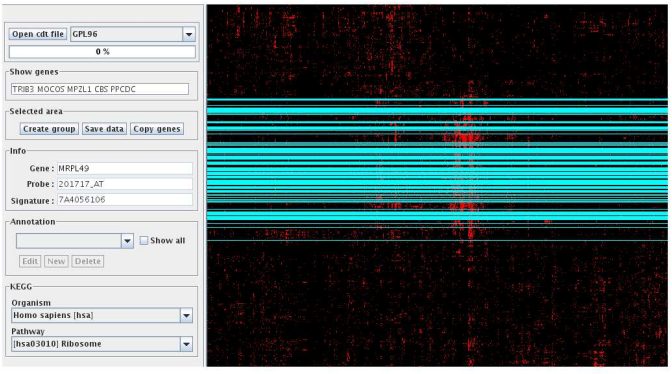
A


Open cdt file
GPL96
0 %
Show genes
TRIB3 MOCOS MPZL1 CBS PPCDC
Selected area
Create group
Save data
Copy genes
Info
Gene : REL
Probe : 206036_S_AT
Signature : 03A3B8CD4
Annotation
Show all
Edit
New
Delete
KEGG
Organism
Homo sapiens [hsa]
Pathway
[hsa00970] Aminoacyl-tRNA biosynthe...

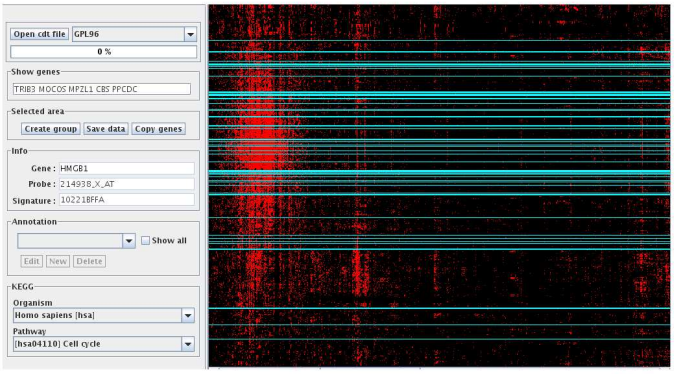
B


Open cdt file
GPL96
0 %
Show genes
TRIB3 MOCOS MPZL1 CBS PPCDC
Selected area
Create group
Save data
Copy genes
Info
Gene : SUPT4H1
Probe : 201483_S_AT
Signature : E3273E73F
Annotation
Show all
Edit
New
Delete
KEGG
Organism
Homo sapiens [hsa]
Pathway
[hsa00190] Oxidative phosphorylation

C


Open cdt file
GPL96
0 %
Show genes
TRIB3 MOCOS MPZL1 CBS PPCDC
Selected area
Create group
Save data
Copy genes
Info
Gene : MRPL49
Probe : 201717_AT
Signature : 7A4056106
Annotation
Show all
Edit
New
Delete
KEGG
Organism
Homo sapiens [hsa]
Pathway
[hsa03010] Ribosome

D


Open cdt file
GPL96
0 %
Show genes
TRIB3 MOCOS MPZL1 CBS PPCDC
Selected area
Create group
Save data
Copy genes
Info
Gene : HMGB1
Probe : 214938_X_AT
Signature : 10221BFFA
Annotation
Show all
Edit
New
Delete
KEGG
Organism
Homo sapiens [hsa]
Pathway
[hsa04110] Cell cycle

Plateform ID	Plateform Manufacturer	Organism	Number of Probes	Title
GPL91	Affymetrix	Homo sapiens	12651	Affymetrix GeneChip Human Genome U95 Version [1 or 2] Set HG-U95A
GPL92	Affymetrix	Homo sapiens	12620	Affymetrix GeneChip Human Genome U95 Set HG-U95B
GPL93	Affymetrix	Homo sapiens	12646	Affymetrix GeneChip Human Genome U95 Set HG-U95C
GPL570	Affymetrix	Homo sapiens	54681	Affymetrix GeneChip Human Genome U133 Plus 2.0 Array
GPL571	Affymetrix	Homo sapiens	22277	Affymetrix GeneChip Human Genome U133A 2.0 Array
GPL80	Affymetrix	Homo sapiens	7129	Affymetrix GeneChip Human Full Length Array HuGeneFL
GPL97	Affymetrix	Homo sapiens	22645	Affymetrix GeneChip Human Genome U133 Array Set HG-U133B
GPL98	Affymetrix	Homo sapiens	8934	Affymetrix GeneChip Human 35K SubA Array Hu35k-A
GPL99	Affymetrix	Homo sapiens	8924	Affymetrix GeneChip Human 35K SubB Array Hu35k-B
GPL94	Affymetrix	Homo sapiens	12644	Affymetrix GeneChip Human Genome U95 Set HG-U95D
GPL95	Affymetrix	Homo sapiens	12639	Affymetrix GeneChip Human Genome U95 Set HG-U95E
GPL201	Affymetrix	Homo sapiens	8793	Affymetrix GeneChip Human HG-Focus Target Array
GPL101	Affymetrix	Homo sapiens	8928	Affymetrix GeneChip Human 35K SubD Array Hu35k-D
GPL100	Affymetrix	Homo sapiens	17856	Affymetrix GeneChip Human 35K SubC Array Hu35k-C
GPL96	Affymetrix	Homo sapiens	22283	Affymetrix GeneChip Human Genome U133 Array Set HG-U133A
GPL3921	Affymetrix	Homo sapiens	22277	Affymetrix GeneChip HT-HG_U133A Array
GPL75	Affymetrix	Mus musculus	6584	Affymetrix GeneChip Murine 11K SubA Array Mu11K-A
GPL32	Affymetrix	Mus musculus	12654	Affymetrix GeneChip Murine Genome U74A Version 1
GPL339	Affymetrix	Mus musculus	22690	Affymetrix GeneChip Mouse Expression Array 430A and Mouse Genome 430A 2.0 Array
GPL340	Affymetrix	Mus musculus	22575	Affymetrix GeneChip Mouse Expression Array 430B
GPL76	Affymetrix	Mus musculus	6595	Affymetrix GeneChip Murine 11K SubB Array Mu11K-B
GPL1261	Affymetrix	Mus musculus	45101	Affymetrix GeneChip Mouse Genome 430 2.0 Array
GPL5722	Affymetrix	Mus musculus	4019	Affymetrix Mus musculus 4K chromosome 11 and 16 CustomExpress Array
GPL83	Affymetrix	Mus musculus	11934	Affymetrix GeneChip Murine Genome U74 Version 2 Set MG-U74C
GPL82	Affymetrix	Mus musculus	12477	Affymetrix GeneChip Murine Genome U74 Version 2 Set MG-U74B
GPL81	Affymetrix	Mus musculus	12488	Affymetrix GeneChip Murine Genome U74 Version 2 Set MG-U74A
GPL1355	Affymetrix	Rattus norvegicus	31099	Affymetrix GeneChip Rat Genome 230 2.0 Array
GPL342	Affymetrix	Rattus norvegicus	15333	Affymetrix GeneChip Rat Expression Set 230 Array RAE230B
GPL85	Affymetrix	Rattus norvegicus	8799	Affymetrix GeneChip Rat Genome U34 Array Set RG-U34A
GPL341	Affymetrix	Rattus norvegicus	15923	Affymetrix GeneChip Rat Expression Set 230 Array RAE230A
GPL88	Affymetrix	Rattus norvegicus	1322	Affymetrix GeneChip Rat Neurobiology U34 Array RN-U34
GPL86	Affymetrix	Rattus norvegicus	8791	Affymetrix GeneChip Rat Genome U34 Array Set RG-U34B
GPL87	Affymetrix	Rattus norvegicus	8789	Affymetrix GeneChip Rat Genome U34 Array Set RG-U34C

Table S2.

Informations related to experiments (n = 1,484) that were analyzed using the DBF-MCL algorithm. All Informations were obtained from the GEO website.

Ce fichier n'est pas reproduit compte tenu du nombre important de pages qu'il représente.

Table S3.

This matrix summarizes the results obtained using the “ESR1 & GATA3 & FOXA1” query. Rows correspond to genes and columns to TS. The presence of a given gene in a given TS is indicated by 1 (default 0).

Ce fichier n'est pas reproduit compte tenu du nombre important de pages qu'il représente.

Gene Expression Profiles Characterize Inflammation Stages in the Acute Lung Injury in Mice

Isabelle Lesur, **Julien Textoris**, Béatrice Loriod, Cécile Courbon, Stéphane Garcia, Marc Leone, Catherine Nguyen

PLoS ONE 5(7): e11485

L'atteinte inflammatoire aiguë du parenchyme pulmonaire est une complication grave de diverses pathologies, notamment infectieuses. Malgré la ventilation mécanique, la mortalité des patients est élevée. La distinction entre une atteinte inflammatoire d'origine infectieuse ou non est difficile, d'autant que l'inflammation du parenchyme pulmonaire favorise la survenue de surinfections et d'un syndrome de défaillance multi-viscérale.

Nous proposons dans cette étude de décrire la modulation du transcriptome pulmonaire en réponse à une agression d'origine non infectieuse. Pour cela, nous avons mis au point un modèle murin d'agression inflammatoire pure du poumon par injection d'acide oléique dans la veine caudale. Les souris ont été sacrifiées à différents temps dans les 24 h après l'injection d'acide oléique. Le choix des temps d'analyse est issu d'une analyse préliminaire de l'expression de certains marqueurs pro- ou anti-inflammatoires par RT-PCR (pics d'expression du TNF, de l'IL4, l'IL6 et l'IL10). L'expression du transcriptome pulmonaire a été analysée par une puce cDNA/nylon de 9 900 sondes. Les gènes modulés de manière précoce (1 h – 1 h 30) sont impliqués dans la réponse immune pro-inflammatoire. Les gènes modulés entre 3 h et 4 h sont impliqués dans l'adhésion inter-cellulaire et le remodelage du cytosquelette. Les gènes modulés plus tardivement (18 h – 24 h) révèlent une modulation de la transcription, de la synthèse des protéines et du métabolisme des lipides. L'évolution au cours du temps de l'expression des cytokines pro- et anti-inflammatoires au sein du parenchyme pulmonaire a été confirmée par RT-PCR.

Ce travail décrit l'évolution de l'expression des gènes au sein du parenchyme pulmonaire dans les 24 h qui suivent une agression inflammatoire. Ces résultats préliminaires posent les bases de l'analyse du transcriptome pulmonaire lors d'une inflammation d'origine infectieuse ou non. Cette étude est suivie d'une seconde analyse réalisée à partir des prélèvements sanguins et analysant la réponse des cellules mononuclées dans le même

modèle d'injection d'acide oléique. Plusieurs voies métaboliques sont activées de manière concomitantes dans le sang et le poumon (réponse immune et inflammatoire, perturbation du métabolisme des lipides). D'autres comme la réparation de l'ADN ou l'autophagie ne sont activées que dans le sang. Enfin, les gènes impliqués dans le contrôle de la transcription sont activés plus précocement dans le sang que dans le poumon. Ces résultats montrent que la réponse inflammatoire à l'injection d'acide oléique est partiellement compartimentalisé. (Cette étude est en cours de révision (R2) à PlosOne et est présentée en annexe 1).

OPEN ACCESS Freely available online

PLoS one

Gene Expression Profiles Characterize Inflammation Stages in the Acute Lung Injury in Mice

Isabelle Lesur[1], Julien Textoris[1,2], Béatrice Loriod[1], Cécile Courbon[3], Stéphane Garcia[4], Marc Leone[2], Catherine Nguyen[1]*

1 Inserm U928, TAGC, Parc scientifique de Luminy, Université de la méditerranée, Marseille, France, **2** Service d'anesthésie et de réanimation, hôpital Nord, AP-HM, Université de la méditerranée, Marseille, France, **3** Service d'anesthésie et de réanimation, hôpital LaTimone, AP-HM, Université de la méditerranée, Marseille, France, **4** Faculté de Médecine – Secteur Nord, labo transfert d'oncologie, Université de la méditerranée, Marseille, France

Abstract

Acute Lung Injury (ALI) carries about 50 percent mortality and is frequently associated with an infection (sepsis). Life-support treatment with mechanical ventilation rescues many patients, although superimposed infection or multiple organ failure can result in death. The outcome of a patient developing sepsis depends on two factors: the infection and the pre-existing inflammation. In this study, we described each stage of the inflammation process using a transcriptional approach and an animal model. Female C57BL6/J mice received an intravenous oleic acid injection to induce an acute lung injury (ALI). Lung expression patterns were analyzed using a 9900 cDNA mouse microarray (MUSV29K). Our gene-expression analysis revealed marked changes in the immune and inflammatory response metabolic pathways, notably lipid metabolism and transcription. The early stage (1 hour–1.5 hours) is characterized by a pro-inflammatory immune response. Later (3 hours–4 hours), the immune cells migrate into inflamed tissues through interaction with vascular endothelial cells. Finally, at late stages of lung inflammation (18 hours–24 hours), metabolism is deeply disturbed. Highly expressed pro-inflammatory cytokines activate transcription of many genes and lipid metabolism. In this study, we described a global overview of critical events occurring during lung inflammation which is essential to understand infectious pathologies such as sepsis where inflammation and infection are intertwined. Based on these data, it becomes possible to isolate the impact of a pathogen at the transcriptional level from the global gene expression modifications resulting from the infection associated with the inflammation.

Citation: Lesur I, Textoris J, Loriod B, Courbon C, Garcia S, et al. (2010) Gene Expression Profiles Characterize Inflammation Stages in the Acute Lung Injury in Mice. PLoS ONE 5(7): e11485. doi:10.1371/journal.pone.0011485

Editor: Carol Feghali-Bostwick, University of Pittsburgh, United States of America

Received December 9, 2009; **Accepted** June 8, 2010; **Published** July 8, 2010

Funding: This work was supported by grants from Conseil Régional Provence Alpes Côte d'Azur, Société Française d'Anesthésie et de Réanimation, and le Ministere de la santé (Programme Hospitalier de Recherche Clinique, PHRC 2005). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: nguyen@tagc.univ-mrs.fr

Introduction

Acute lung injury (ALI) is a diffuse lung injury which is characterized by a widespread capillary leakage leading to hypoxemia and low lung compliance. ALI is caused by either a direct (e.g. pneumonia) or indirect (e.g. pancreatitis) injury of the lung. Any local (e.g. pneumonia) or systemic inflammation (e.g. pancreatitis) can lead to a critical alteration of the lung function. The Acute Respiratory Distress Syndrome (ARDS) is the last stage of this acute inflammatory process and still carries a high mortality rate (40–50%) [1]. Sepsis is a major cause of ARDS either by direct alteration of the lung or indirectly through the Systemic Inflammatory Response Syndrome (SIRS) associated with severe sepsis. Inflammation is part of the defense mechanisms of innate immunity, occurring after tissue injury. At the site of inflammation, a cascade of mediators such as cytokines initiates activation of inflammatory cells (early-inflammatory phase). Then, white blood cells migrate through the wall of blood vessels and infiltrate the surrounding tissues. ARDS frequently occurs in a context of severe sepsis in which inflammation and infection interplay. Then, it is unclear to delineate the pathways related to inflammation or infection. In sepsis, the production of both pro- and anti-inflammatory cytokines in sepsis has been widely studied [2]. While pro-inflammatory cytokines are necessary for initiating an effective inflammatory process against infection, anti-inflammatory cytokines seem to be a prerequisite for controlling and down-regulating the inflammatory response leading to a depression of the immune system of patients [3]. Human immune responses to sepsis are mediated mainly by the primary pro-inflammatory cytokines. The timing of cytokine release and the balance between pro- and anti-inflammatory mediators seems to be associated with the severity of sepsis [4–5]. An excessive production may induce deleterious effects [5]. To this purpose, deciphering gene expression profiles of either infection or inflammation alone seems crucial.

One way to do so consists on studying transcriptional genes expression profiles at several inflammation stages (with or without sepsis). Genome-wide gene expression profiling using microarray technology has been applied successfully to the study of human disease pathogenesis. Examples include the discovery of new cancer subtypes with different prognosis and response to therapies [6], or new hypotheses of disease pathogenesis in ARDS [7]. Gene expression analysis by microarray studies of genes clusters that have similar expression changes over time allows the definition of

functionally meaningful expression patterns. This approach has been successfully used to study either ARDS, using either a whole organ or cultured cells or in severe sepsis studies where inflammation was associated with infection [8].

In humans with ARDS, time-course studies using blood samples made it possible to obtain successful results using a transcriptional approach. Wang *et al.* studied global gene expression profiling in 8 human blood samples and identified potential candidate genes that can be used as biomarkers for ARDS [9]. They reported for instance the role of peptidase inhibitor 3 (Serpina1c/PI3) encoding Elafin which has antimicrobial and anti-inflammatory activities. One limitation in using human samples is the heterogeneity of the samples with patients from different gender, age, injury, ethnic origin and genetic backgrounds.

Previous transcriptional studies performed to study ARDS used various experimental techniques, various species and various types of infection. The common point between these studies is the sequence of events: first, an infection is associated with an inflammation and then, samples are studied using a transcriptional approach. At this point, since gene expression patterns depend on both inflammation and infection, it is not possible to identify the impact of either inflammation or infection on gene expression. A similar infective agent leads to different outcomes depending on the pre-existent inflammation stage of each patient [4]. For this reason, we decided to set up a mouse model to study ALI without any infection. Inflammation induced by intravenous administration of Oleic Acid (OA) resembles ARDS in many morphological, histological, and physiological respects [10]. OA-induced ALI is consistently associated with acute respiratory failure characterized by hypoxemia and reduced lung compliance due to alveolar damage, intra-alveolar hemorrhage, and leakage of proteinaceous fluid into the air space [11]. The progression of OA-induced injury is much shorter than that of human ARDS. The acute phase of human ARDS develops over 1–7 days. However, it is also characterized by extensive damage to the alveolar epithelium, hemorrhage, and pulmonary edema. Despite some limitations, the OA-ALI remains a relevant model for investigating the effects of ALI. Indeed, no other model has been used as extensively to evaluate the risks and benefits of various treatment strategies before applying them to patients with ARDS [10].

Our objective was to identify some biologically relevant process important in ALI by studying the molecular interactions between OA and inflammation in a mouse model. The modest success of microarray approaches to identify novel potential therapeutic targets may be due to a lack of comprehension to how changes in gene expression occur through time. For this reason, we decided to investigate lung gene expression patterns in female mice developing inflammation without any infection using a 9900 cDNA mouse microarray over a 24 hours time-course. By looking at changes in gene expression over time, we developed a global strategy to identify novel pathways that could represent novel targets for therapy.

Results

Outcome of mice and definition of pro- and anti-inflammatory phases by real-time PCR

Following the injection of OA, we noticed a transitional prostration of the mice which started 1.5 hours after injection and lasted for several hours. All mice recovered; the survival rate was 100% after 24 hours incubation. To evaluate for the incubation period necessary to induce lung inflammation, we considered a wide time-course ranging from 1 hour to 24 hours after OA injection even if previously published studies considered much shorter incubation times [12–13]. Prior to large-scale gene expression analysis, we quantified expression level of one pro-inflammatory cytokine (Tnfα), one cytokine with pro- and anti-inflammatory activities (Il6), and two anti-inflammatory cytokines (Il10 and Il4) at 1 hour, 1.5 hours, 3 hours, 4 hours, 18 hours and 24 hours after OA and physiological serum injection (figure 1 A). Expression levels of Tnfα, Il6, Il10 and Il4 peaked respectively at 1.5 hours, 3 hours, 4 hours and 3 hours-4 hours after OA injection. Thus, inflammatory response ranged from 60 minutes to several hours following OA injection. Real-time PCR allowed us to identify a pro-inflammatory expression peak 1.5 hours after OA injection (figure 1 B), a transition phase 3 hours to 4 hours after OA injection (figure 1 C) and an anti-inflammatory phase 4 hours after OA injection (figure 1 D, E).

Identification of the inflammation timing following OA injection using lung histology

Compared with physiological serum controls, OA induced a transient increase in histological lung damage 1.5 hours after injection (figure 2). Histologically, 1.5 hours after OA injection, we found a neutrophil accumulation in the alveolar wall associated with edema. Lung histological modifications are thus correlated with the transcriptional up-regulation of pro-inflammatory mediators (both by microarray and RT-PCR). These results strengthened our real-time PCR data. There were no more histological lung lesions 3 hours after OA injection.

Microarray analysis discriminated between early- and late-inflammatory stages according to the time-course after injection

To focus on transcriptional changes associated with the inflammatory process in lung, we searched for a set of genes that discriminated between mice at different incubation times. To this purpose, we used the multi-class SAM procedure [14], and we applied a false discovery rate of 5% (figure 3 and see supplementary data file S1). The analysis yielded a set of 1000 genes (see supplementary data file S2). By performing a hierarchical clustering of the samples (microarrays), we identified 3 clusters (see supplementary data file S3). Samples at 1 hour and 1.5 hours clustered together. The same way, samples at 3 hours and 4 hours were grouped together. Finally, samples at 18 hours and 24 hours clustered together with the exception of three 18 hours samples which have been grouped with the 3 hours–4 hours cluster as an out-group.

We also performed unsupervised hierarchical clustering on the 1000 genes (figure 4). Interestingly, the expression level of numerous genes was induced specifically either early or late after OA injection while other genes were specifically up-regulated during the transition phase between the early- and late-inflammatory stages (see supplementary data file S4). The Cluster software [15] classified the 1000 significantly differentially expressed genes into 3 main clusters. Genes of cluster Early (365 genes) were over-expressed at early stage of inflammation (1 hour–1.5 hours); their expression level decreased at 3–4 hours and then these genes were under-expressed at late time-points (18 hours–24 hours). Genes of the Intermediate cluster (269 genes) were over-expressed between 1.5 hours and 4 hours compared to the earliest time-point (1 hour) and the late measurements (18 hours –24 hours), which correspond to the transition between early- and late-inflammatory responses. At the latest stage of inflammation, genes from the Late cluster (366 genes) were highly over-expressed. Expression level of these genes increased at 3 hours to 4 hours after OA injection and reached a peak at 18 hours–24 hours.

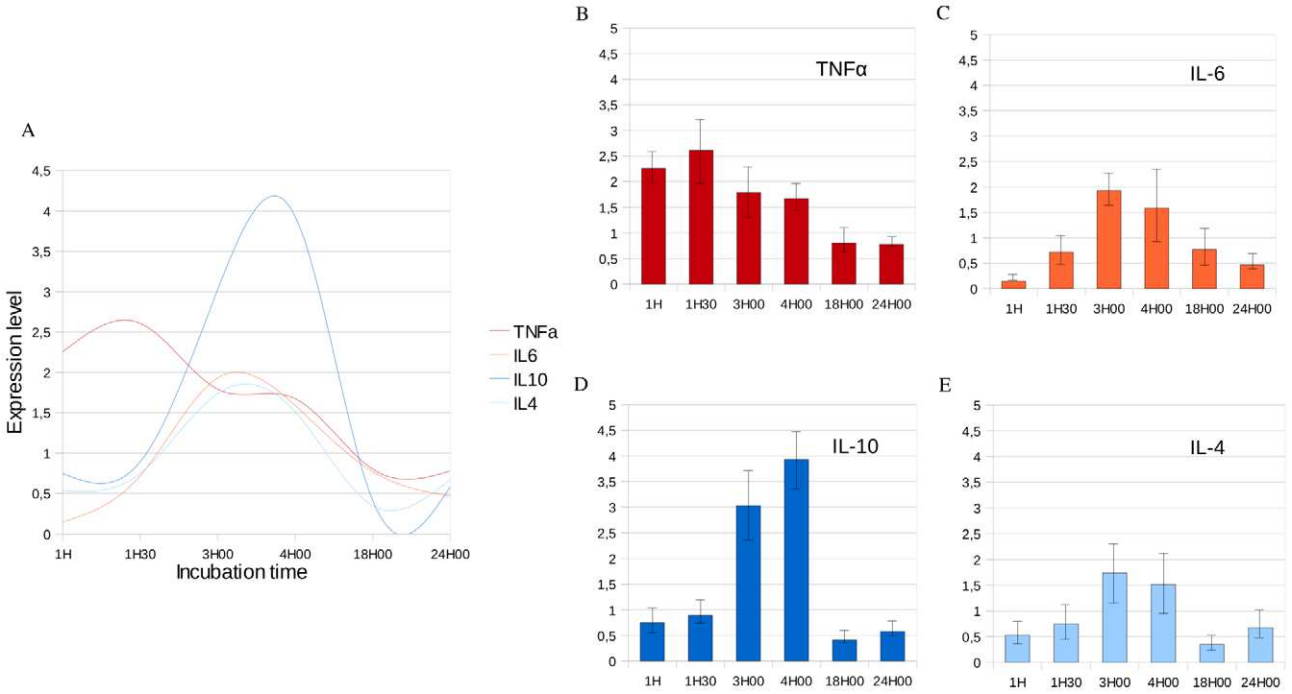


Figure 1. Variation of expression of pro-inflammatory and anti-inflammatory cytokines determined by real-time RT-PCR in lung samples. The X-axis corresponds to the incubation time after OA injection. The Y-axis corresponds to the ratio between the expression level of a cytokine measured after OA injection and the expression level of this same cytokine measured after the same incubation time following physiological serum injection. Measurements are normalized using the βactin housekeeping gene. A- Variations of expression level of Tnfα, Il6, Il10 and Il4 across OA incubation time allowed us to identify pro- and anti-inflammatory phases. B- We identified the pro-inflammatory peak 1H30 after injection of OA which corresponds to the maximum expression level of Tnfα. C- Il6 which is both a pro- and anti-inflammatory cytokine is highly expressed 3H after OA injection. D- The expression level of Il10, an anti-inflammatory cytokine is maximal 4H after injection of OA. E- Il4 is highly expressed 3H to 4H after injection of OA. We observe that the anti-inflammatory response occurs approximately 4H after OA injection and follows the pro-inflammatory response which occurs approximately 1H30 after OA injection.
doi:10.1371/journal.pone.0011485.g001

To analyze functional annotations related to the pulmonary inflammation, we considered biological process Gene Ontology (GO) terms, Kegg pathways and the Biocarta database using DAVID for the 365 genes belonging to the cluster Early, the 269 genes belonging to the cluster Intermediate, and the 366 genes belonging to the cluster Late.

GO terms related to the «Defense response» ($p = 9.9 \times 10^{-3}$) such as «Inflammatory response» ($p = 6.9 \times 10^{-3}$), "cytokine activity" ($p = 1.1 \times 10^{-2}$) and «Immune response» ($p = 6.5 \times 10^{-3}$) were strongly represented in cluster Early (table 1). Similarly, several genes grouped in the cluster Early were found to be involved in Kegg pathways related to immune responses such as

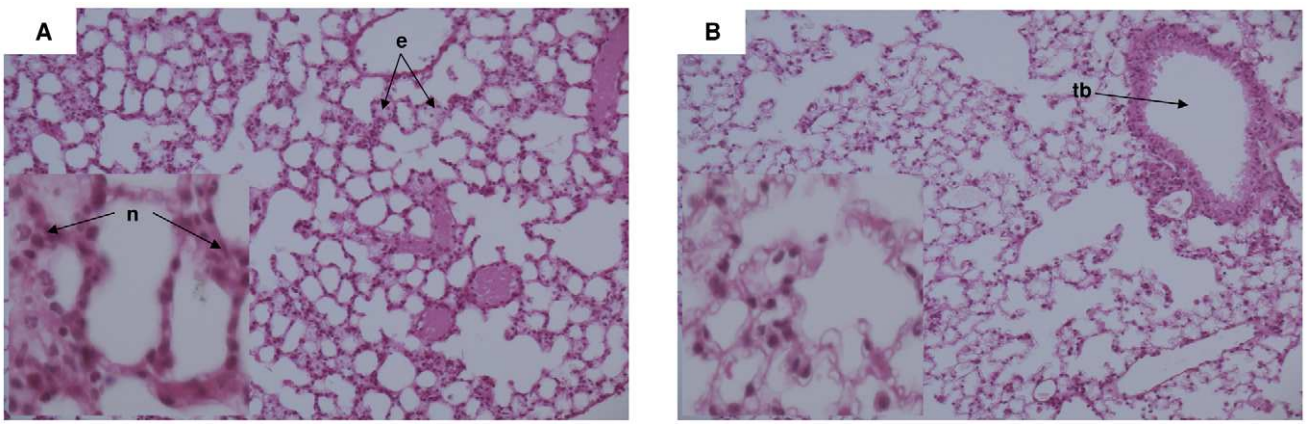


Figure 2. Histological quantification of lung injury following OA injection. Sections are representative of all animals examined per group. A- Section of lung 1H30 after OA injection showing interstitial neutrophils (n) and edema (e). There was no significant fibrin formation within the alveolar spaces – B- Section of lung 3H00 after OA injection showing absence of edema or fibrin. There was no more significant neutrophil infiltrate - (tb) Terminal bronchiole. (H&E x100, cartridges X400).
doi:10.1371/journal.pone.0011485.g002

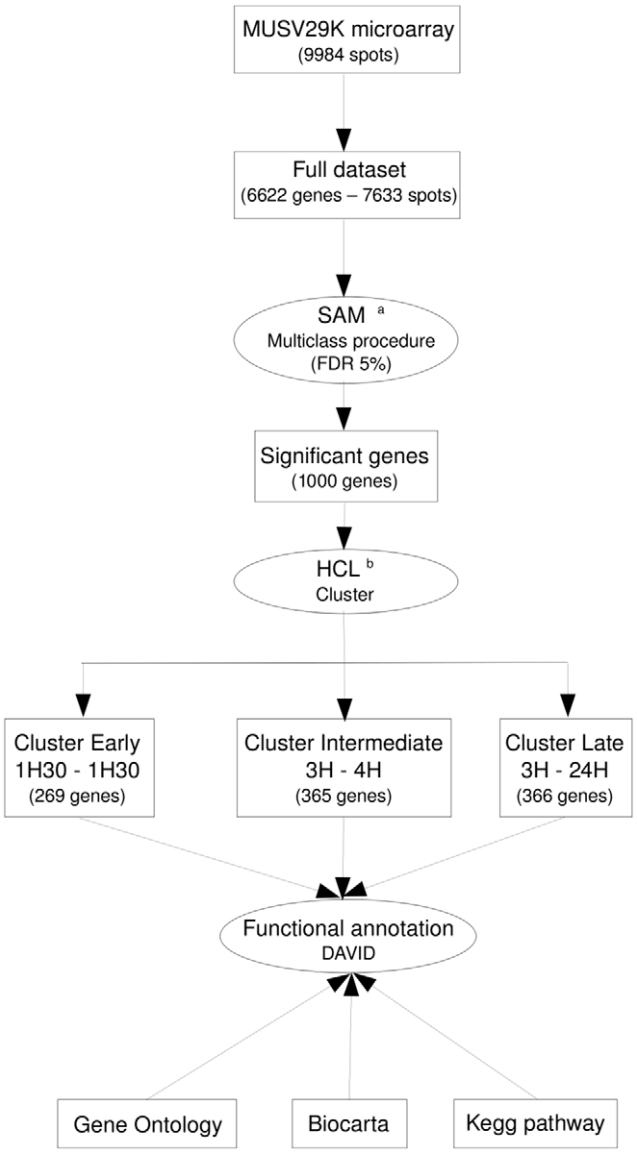


Figure 3. Schematic outline of data analysis. [a] We compared gene expression between three groups: 1H-1H30/3H-4H/18H-24H. [b] HCL: Hierarchical Clustering. We considered the full dataset (n = 6622 genes) to identify genes significantly differentially expressed between the three classes previously described. The 1000 genes identified with SAM have been clustered using the Cluster software and each cluster was independently functionally annotated using DAVID.
doi:10.1371/journal.pone.0011485.g003

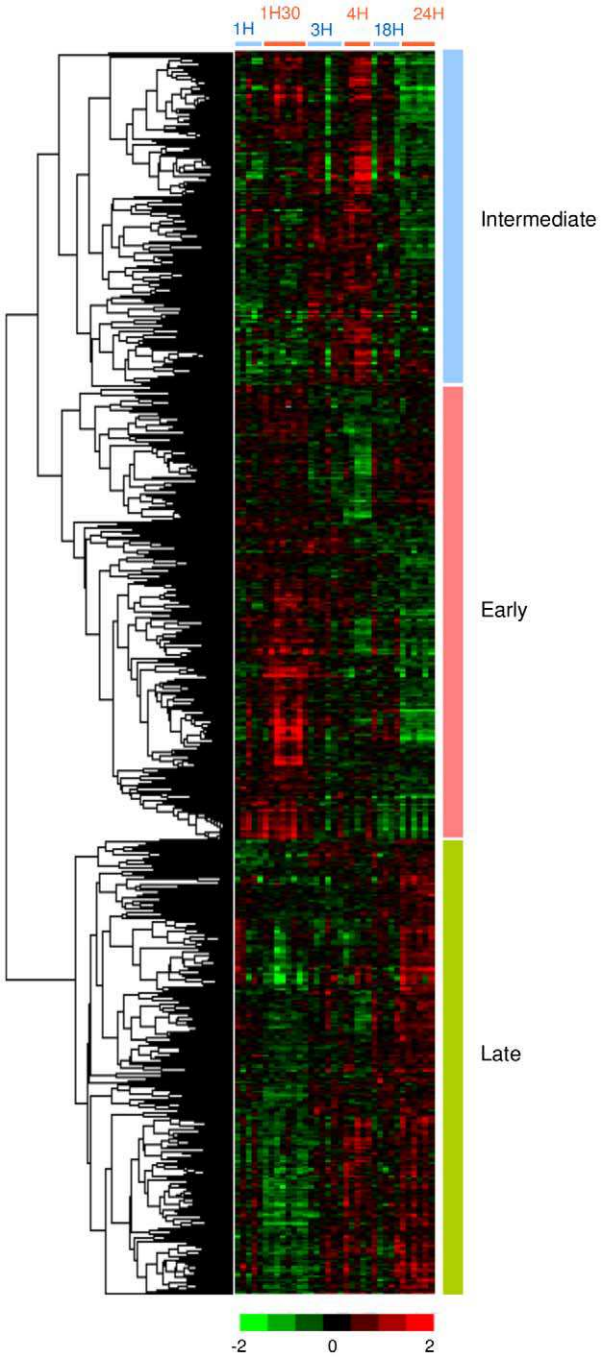


Figure 4. Hierarchical classification of 1000 significant genes differentially expressed at early and late stages of inflammation. This set of genes was extracted from the full dataset (n = 9984) by use of a SAM procedure and a false discovery rate of 5%. Each row represents a gene and each column represents a sample. Red and green indicate expression levels respectively above and below the median. Dendrogram of genes, to the left of the matrix represents overall similarities in gene expression profiles. 3 clusters have been identified: Early, Intermediate and Late.
doi:10.1371/journal.pone.0011485.g004

«MAPK signaling pathway» ($p = 1.4 \times 10^{-2}$). The Functional Annotation Clustering tool from DAVID grouped these terms in a single annotation cluster with the best enrichment score (1.36). Nine genes belonging to the «Hematopoietic cell lineage» pathway ($p = 6.5 \times 10^{-2}$) were present in this cluster. As expected in inflammation response, immune and pro-inflammatory responses are activated. These results validate our experimental strategy, particularly the use of OA to induce a reliable inflammation without any infection.

Most of the genes grouped in the cluster Intermediate were found to be involved in Kegg pathways related to the immune response, notably the spreading of immune cells into inflamed tissues through interaction with vascular endothelial cells, such as «Leukocyte Transendothelial migration» ($p = 7.8 \times 10^{-3}$) or «Adherens junction» ($p = 3.5 \times 10^{-2}$) (table 2). This was further supported by the analysis of Biocarta Pathway, which pointed out several pathways such as «Adhesion Molecules on Lymphocyte» ($p = 8.5 \times 10^{-3}$) and «Monocyte and its Surface Molecules» ($p = 1.8 \times 10^{-2}$). These pathways were clustered together by the Functional Annotation Clustering tool

Table 1. Functional annotation and enrichment of the 365 genes belonging to cluster Early.

Functional classification			
Database	*Term*	*Count*	*P-Value*
Gene Ontology	Immune response	23	6.50E-03
Gene Ontology	Inflammatory response	14	6.90E-03
Gene Ontology	Defense response	18	9.90E-03
Gene Ontology	Cytokine activity	13	1.10E-02
Kegg pathway	MAPK signaling pathway	16	1.40E-02
Gene Ontology	Response to wounding	16	1.90E-02
Gene Ontology	Immune system process	30	2.30E-02
Gene Ontology	Protein kinase cascade	14	3.10E-02
Gene Ontology	Positive regulation of kinase activity	8	3.30E-02
Gene Ontology	Chemokine activity	6	5.00E-02
Gene Ontology	MAP kinase tyrosine/serine/threonine phosphatase activity	3	5.10E-02
Gene Ontology	MAP kinase phosphatase activity	3	5.10E-02
Gene Ontology	Chemokine receptor binding	6	5.60E-02
Kegg pathway	Hematopoietic cell lineage	9	6.50E-02
Gene Ontology	Interleukin-1 receptor binding	3	6.90E-02
Gene Ontology	MAPKKK cascade	7	7.40E-02
Annotation Cluster 1	**Enrichment score: 1.36**		
Database	*Term*	*Count*	*P-Value*
Gene ontology	Immune response	23	6.50E-03
Gene ontology	Inflammatory response	14	6.90E-03
Gene ontology	Defense response	18	9.90E-03
Gene ontology	Cytokine activity	13	1.10E-02
Gene ontology	Response to wounding	16	1.90E-02
Gene ontology	Immune system process	30	2.30E-02
Gene ontology	Response to external stimulus	19	3.30E-02
Gene ontology	Chemokine activity	6	5.00E-02
Gene ontology	Chemokine receptor binding	6	5.60E-02

doi:10.1371/journal.pone.0011485.t001

Table 2. Functional annotation and enrichment of the 269 genes belonging to cluster Intermediate.

Functional classification			
Database	*Term*	*Count*	*P-Value*
Kegg pathway	Leukocyte transendothelial migration	8	7.80E-03
Biocarta	Adhesion molecules on lymphocyte	4	8.50E-03
Biocarta	Neutrophil and its surface molecules	4	8.50E-03
Biocarta	Monocyte and its surface molecules	4	1.80E-02
Kegg pathway	Adherens junction	6	3.50E-02
Kegg pathway	Hematopoietic cell lineage	7	4.20E-02
Annotation Cluster 1	**Enrichment score: 1.16**		
Database	*Term*	*Count*	*P-Value*
Biocarta	Adhesion molecules on lymphocyte	4	8.50E-03
Biocarta	Neutrophil and its surface molecules	4	8.50E-03
Biocarta	Monocyte and its surface molecules	4	1.80E-02
Gene Ontology	Wound healing	5	9.40E-02

doi:10.1371/journal.pone.0011485.t002

from DAVID in a single cluster with an enrichment score of 1.16. In addition, 5 genes involved in differentiation of T and B lymphocytes cells were found to be involved in the «Hematopoietic cell lineage» pathway ($p = 4.2 \times 10^{-2}$). Surprisingly, genes involved in the spreading of inflammation were grouped together. Usually, this process can not be clearly visualized because of the very short period of time between activation of the inflammatory and immune and anti-inflammatory response.

Functional annotation analysis of genes belonging to the cluster Late showed an over-representation of GO terms related to transcription, metabolism and in particular lipid metabolism (figure 5 and supplementary table S1). The Functional Annotation Clustering tool from DAVID identified 3 groups of genes with the highest enrichment scores. The first one (enrichment score = 2.07) showed an over-representation of GO terms related to «Transcription» ($p = 1.4 \times 10^{-3}$), such as «Regulation of transcription» ($p = 1.1 \times 10^{-3}$) or «Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process» ($p = 7.7 \times 10^{-4}$). Most of the GO terms grouped in the second annotation cluster (enrichment score = 1.96) were found to be involved in metabolism, such as «Metabolic process» ($p = 2.2 \times 10^{-3}$) or «Cellular metabolic process» ($p = 2.3 \times 10^{-3}$). The third annotation cluster (enrichment score = 1.44) included GO terms related to lipid metabolism such as «Cholesterol metabolic process» ($p = 9.3 \times 10^{-3}$) or «Cellular lipid metabolic process» ($p = 5.2 \times 10^{-2}$). Increased expression level of genes involved in transcription is a consequence of the metabolism which is highly perturbed. Multiple alterations in lipid and lipoprotein metabolism have been previously described associated with infection and inflammation [16]. Here, we observed that disturbance of genes expression levels related to lipid metabolism is associated only with inflammation.

We also considered genes of interest for clinical studies such as Smad7 [17]. In our study, Smad7 belongs to the set of significantly differentially expressed genes across time and it belongs to cluster Late. Its expression level increased between 3 hours and 4 hours after OA injection to reach a maximum expression level 24 hours after injection. We also considered the angiotensin-converting enzyme gene (Ace) [18–19]. In our study, Ace expression level did not significantly change over time. Finally, we considered transcription of genes involved in surfactant lipid biosynthesis and surfactant-associated proteins. In a previous study, authors quantified expression levels of 5 genes involved in the surfactant lipid biosynthesis and 5 genes involved in the surfactant-associated protein biosynthesis [20]. Expression levels of these genes decreased significantly in mice developing ALI. Out of these 10 genes, 4 are present on MUSV29K: Slc34a, Fasn, Sftpd and Napsa. In our study, Slc34a, involved in surfactant lipid biosynthesis, is the only gene whose expression level changed significantly over time. It belongs to the cluster Late and its expression level increased over time after OA injection to reach a peak at 24 hours incubation.

Overall, our gene expression analysis revealed marked changes in pathways involved in the immune and inflammatory responses, and also lipid metabolism and transcription. First, the pro-inflammatory phase is activated with high expression level of genes involved in inflammation and immune response. Then, inflammation is spreading into inflamed tissues. And, later, transcription is activated and lipid metabolism seems highly disturbed.

Discussion

Data summary

In this study, we have searched for genes and physiological pathways potentially involved in ALI. To this purpose, we analyzed differentially expressed genes in lung from C57BL/6J mice at different stages of inflammation. We found that injection of OA deeply alters gene expression. In particular, we identified 3 clusters of genes specifically highly expressed either at early time-points (early-inflammatory stage), at late time-points (late-inflammatory stage), or at the time-points when the shift between early- and late-inflammatory stages occurs. In order to identify genes whose changes in expression levels are associated with lung inflammation, we performed a three class SAM procedure with a stringent false discovery rate of 5%. The procedure was conducted on the whole dataset. Because the technical procedure may affect the gene expression level, we adjusted the measurements of mice treated with OA to that treated with physiological serum. Thus, we identified 1000 genes differentially expressed between OA mice and physiological serum mice.

Limits of the OA model

In this study, our first aim was to define a model of lung inflammation in mouse using OA. Despite its widespread use, the OA-induced ALI model has not been standardized. One to 2 ml of OA generates ALI when injected as a single bolus directly into the pulmonary circulation to an animal weighing 25 Kg [10]. However, both smaller and larger doses have also been used. At other times, the injection has been given in fractions, after sonication in saline, after OA dissolution in ethanol, or as continuous infusion. Each of these variations may alter the severity or extensiveness of the injury, resulting in different physiopathologic consequences. For example, Zhou *et al.* determined that an intravenous dose of 0.15 μL/g body weight OA in mice resulted in an approximate 24-hour survival rate of 75% [12]. Sixty minutes after OA injection, they observed severe alveolar damage with the development of alveolar edema, increased-permeability and abnormalities in oxygenation. Using doses of OA that were up to approximately 1.5 times the dose used by Zhou *et al.*, Ulrich *et al.* reported similar histopathological findings 1 hour after OA injection in mice [13]. A dose-related difference in mortality was observed in animals treated with 0.2 and 0.4 μL/g of OA administered through the tail vein, respectively. The higher dose was associated with 100% mortality, 30 to 40 minutes following OA administration.

Previous studies showed that a time-course ranging from 40 to 90 minutes was required to detect an inflammatory response [12]. The length of the incubation period required to detect a pro-inflammatory profile (expression levels of pro- and anti-inflammatory cytokine or lung histology) depended on the OA dosage. As mechanical ventilation induces by itself an inflammation [21], one can hypothesize that models using mechanical ventilation increased the inflammatory response. Given the contradictory results previously published regarding the kinetics of inflammation in mice, we quantified mRNA levels of pro- and anti-inflammatory cytokines before performing high-throughput gene expression analysis. Cytokines play a critical role as signaling molecules that initiate, amplify, and perpetuate inflammatory responses on a local and systemic basis.

Timing of inflammation

We quantified expression levels of 4 cytokines 1 hour, 1.5 hours, 3 hours, 4 hours, 18 hours and 24 hours after OA incubation. Tnfα, Il6, Il10 and Il4 are commonly used to specifically study pro- and anti-inflammatory phases (figure 1) [22].

Tnfα is an early response cytokine. It plays an important role in immunity and inflammation, in the control of cell proliferation, differentiation and programmed cell death [23]. We found that, in

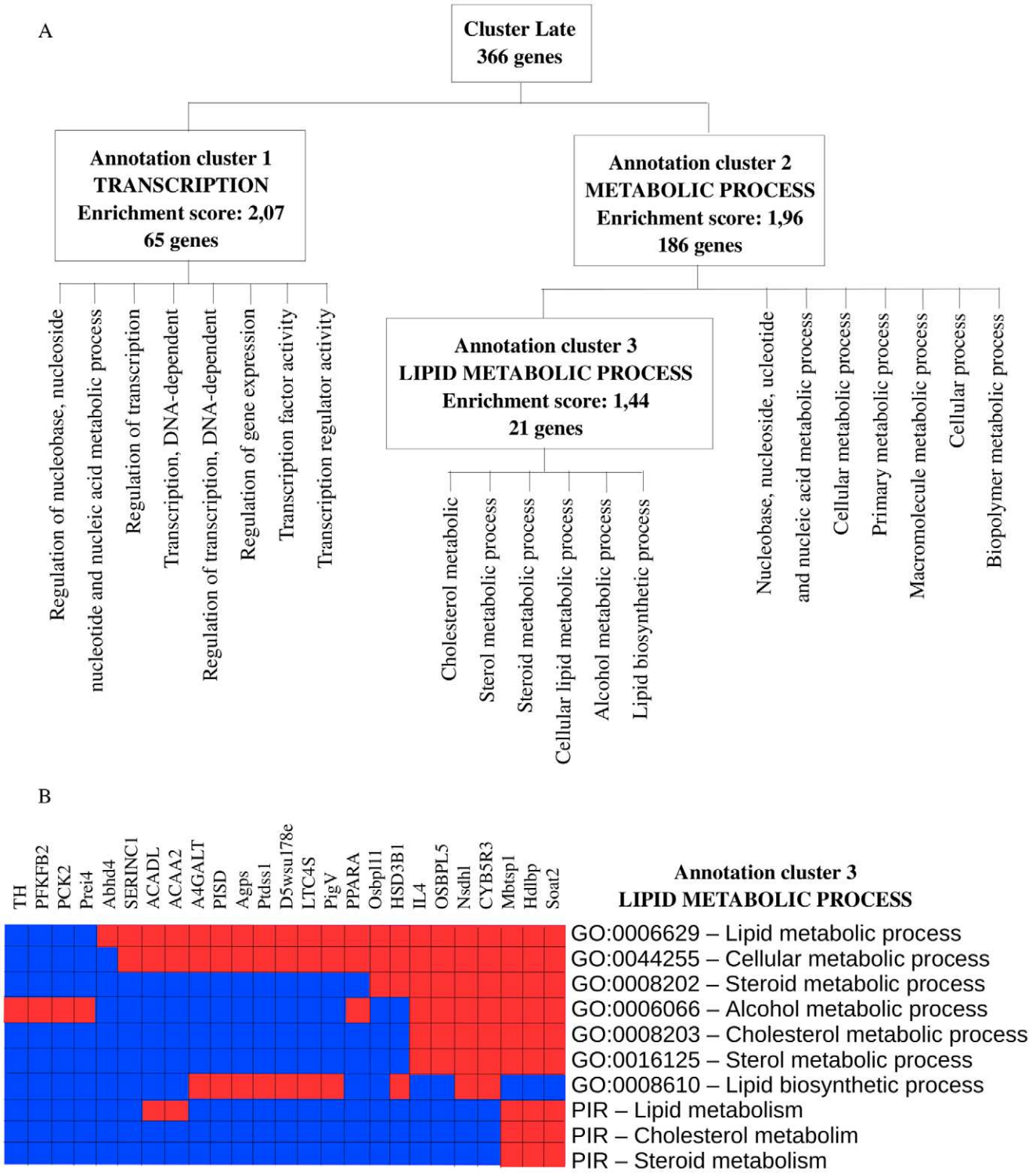


Figure 5. Functional annotation enrichment of the 366 genes belonging to cluster Late. Expression levels of genes belonging to cluster Late increases 3H-4H after OA injection and reach a peak after 18H to 24H incubation. A- Distribution of the 366 genes belonging to Cluster Late into the 3 main annotation clusters identified by the functional annotation clustering tool from DAVID: Transcription, Metabolic process and Lipid metabolic process. B- 21 genes in cluster Late belong to the lipid metabolic process. These genes can be assigned to one or several of 7 GO categories related to Lipid metabolism. Three of them also belong to 3 PIR categories (Protein Information resources) related to lipid metabolism. Red means the gene belong to the pathway and Blue means that the gene does not belong to the pathway.
doi:10.1371/journal.pone.0011485.g005

our model without infection, Tnfα is temporarily highly expressed 60 to 90 minutes after OA injection (figure 1 B). This corresponds to the pro-inflammatory response.

Il6 is produced by a wide range of cells including monocytes/macrophages, endothelial cells, fibroblasts, and smooth muscle cells in response to stimulation by endotoxin [24]. Raised levels of Il6 have been described in a number of acute conditions such as burns, major surgery and sepsis [25]. Il6 is a ubiquitous cytokine playing a role on pro- and anti-inflammation. Several experimental models suggest a protective role for Il6 against inflammation [26]. Pro-inflammatory effects of Il6 have been shown in several tumor cell lines [27]. Il6 is also involved in the spreading of pro-inflammatory cells [28]. In our study, expression level of Il6 is maximum 3 hours after OA injection, interconnecting the pro- and anti-inflammatory phases (figure 1 C).

Il10, an anti-inflammatory cytokine [29], inhibits the release of pro-inflammatory cytokines such as Tnfα from monocytes/macrophages, thus preventing subsequent tissue damage [30]. The role of Il10 is essential to counter balance the pro-inflammatory response. In a previous study, Park *et al.* found a disproportionate anti-inflammatory response to the early pro-inflammatory response, which leads to ARDS [31]. In our study, Il10 expression increased from 3 hours and peaked 4 hours after OA injection which corresponds to the anti-inflammatory response to Tnfα (figure 1 D).

Il4 is also an anti-inflammatory cytokine which down-regulates local levels of pro-inflammatory cytokines and chemokines [32]. We found that its production was simultaneous with Il10 expression peak. This finding confirms the timing of the anti-inflammatory phase (figure 1 E).

The quantification of expression levels of 4 pro- and anti-inflammatory cytokines by real-time PCR allowed us to identify two phases during inflammation. We identified the pro-inflammatory phase based on the high expression level of Tnfα 1.5 hours after OA injection. Later, 3 to 4 hours after OA injection, anti-inflammatory cytokines Il10 and Il4 are highly expressed. Meantime, Il6, which has pro- and anti-inflammatory activities, is highly transcribed 3 hours after OA injection. Given the contradictory results previously published, this preliminary step was essential to timely identify pro- and anti-inflammatory phases in our model and then perform large-scale transcriptional study considering each phase of inflammation described.

An early inflammatory and immune response

Once our mouse model of lung inflammation settled, we identified molecular pathways associated with lung injury.

Injection of OA rapidly affects the expression of genes involved in the defense response such as inflammatory and immune responses. Expression level of numerous chemokines such as Cxcl5, Ccl22, Ccl4, Ccl3, Ccl6 or Ccl25 increased rapidly and allowed recruitment of specific leukocyte subpopulations to sites of tissue damage (see supplementary data file S4) [33]. The expression of pro-inflammatory cytokines and chemokines has been previously observed in transcriptional studies of infection in model organisms. Calvano *et al.* showed, in human blood leukocytes, an over-expression of pro-inflammatory chemokines 2 hours to 4 hours after endotoxin administration [34]. The activation of cytokines following inflammation has also been observed in lung by Chinnaiyan *et al.* in a transcriptional study of systemic inflammation in a cecal ligation/puncture model of sepsis in rat [35]. Also, the MAPK signaling pathway, involved in cell proliferation, differentiation and inflammation is activated (see supplementary data file S4) (figure 6). Three regulated groups of MAPK are activated: the extracellular signal-related kinases (ERK) with genes such as Fgf1, Rasa2, Ras, Dusp1, Mknk1, Mknk2 or Rps6ka3; the Jun amino-terminal kinases (Mapk8, Mapk9, Mapk10) with Il1, Map4k1, Pak1, Pak2 and Dusp1; and the p38 proteins with Il1, Map3k7ip1, Map3k7ip2, Dusp1 and Map3k6. Il1b is known to participate in ARDS pathogenesis [36]. The MAPK cascade pathways are critical for the transmission of activated cell surface receptor signals to invoke multiple regulated intracellular processes. We also noticed over-expression of Il15ra which blocks Il15, an inhibitor of Tnfα [37]. Activation of inflammatory and immune responses early after OA injection validates our inflammation model in mouse [23].

An intermediate spreading of inflammation

The transition phase at 3 hours –4 hours which followed the pro-inflammatory phase corresponds to the spreading of inflammation. Among genes whose expression level reaches a maximum at this time-point, we found transmembrane adhesion molecules (Cd44, Pecam1 and Icam1) as well as endothelial cell adhesion molecule (CAM) (Cd34, Icam1, Glycam1 and Pecam1) (see supplementary data file S4) [37–38]. It has been previously shown that CAM adhesion molecules are strongly involved in the inflammatory process by modulating the leukocyte trafficking [38–39]. B and T cell lymphocytes interact with a variety of cells as part of their immune function allowing inflammatory signals to circulate between organs. The interaction of lymphocytes with other cell types like vascular endothelial cells requires transmembrane adhesion molecules. Each of these lymphocyte adhesion molecules interacts with specific ligands expressed on cells like endothelial cells to moderate adhesion between cells. Cd44 is an adhesion molecule with several roles, including interaction of leukocytes with endothelial cells in response to inflammation [40]. Pecam1 mediates the spreading of monocytes and other immune cells into inflamed tissues through interaction with vascular endothelial cells [41]. Expression of intercellular adhesion molecule-1 (Icam1) is increased by inflammatory signals, leading to higher adhesion and permeation of lymphocytes into inflamed tissues. Icam1 is an inducible protein expressed on the surface of endothelial cells. Under physiological conditions, Icam1 is not constitutively expressed, or is expressed at low levels in most tissues. In inflammatory states, it has been shown that expression level of Icam1 is up-regulated by Tnfa [42]. It has been shown that increased expression of Icam1 protects lung against injury by neutralizing antibodies or targeted gene deletion during bacterial sepsis, acute pancreatitis and trauma [43]. As Icam1 deficiency interferes with neutrophil recruitment, these results support the concept of a therapeutic strategy directed against neutrophil migration and activation. During inflammation, leukocytes bind to endothelial CAM and then migrate across the vascular endothelium.

There is a clear link between blood and inflammation as blood carries the inflammation throughout the entire organism. We found, in cluster Intermediate, 9 genes such as Cd34, Cd44, Il7r, Csf1 and Cd38 leading to differentiation of blood cells (see supplementary data file S4). Blood-cell development progresses from an hematopoietic stem cell which can differentiate into several lineages such as T lymphocytes, B lymphocytes, NK cells, erythrocytes or platelets. Cells undergoing this differentiation process express a stage- and lineage-specific set of surface markers. Therefore cellular stages are identified by the specific expression patterns of these genes. Cd34 is an adhesion molecule with a role in early hematopoiesis by mediating the attachment of stem cells to the bone marrow extracellular matrix or directly to stromal cells [44]. Cd44 is involved in lymphocyte activation [40]. Il7r is the receptor of Il7 which is involved in the regulation of lymphopoi-

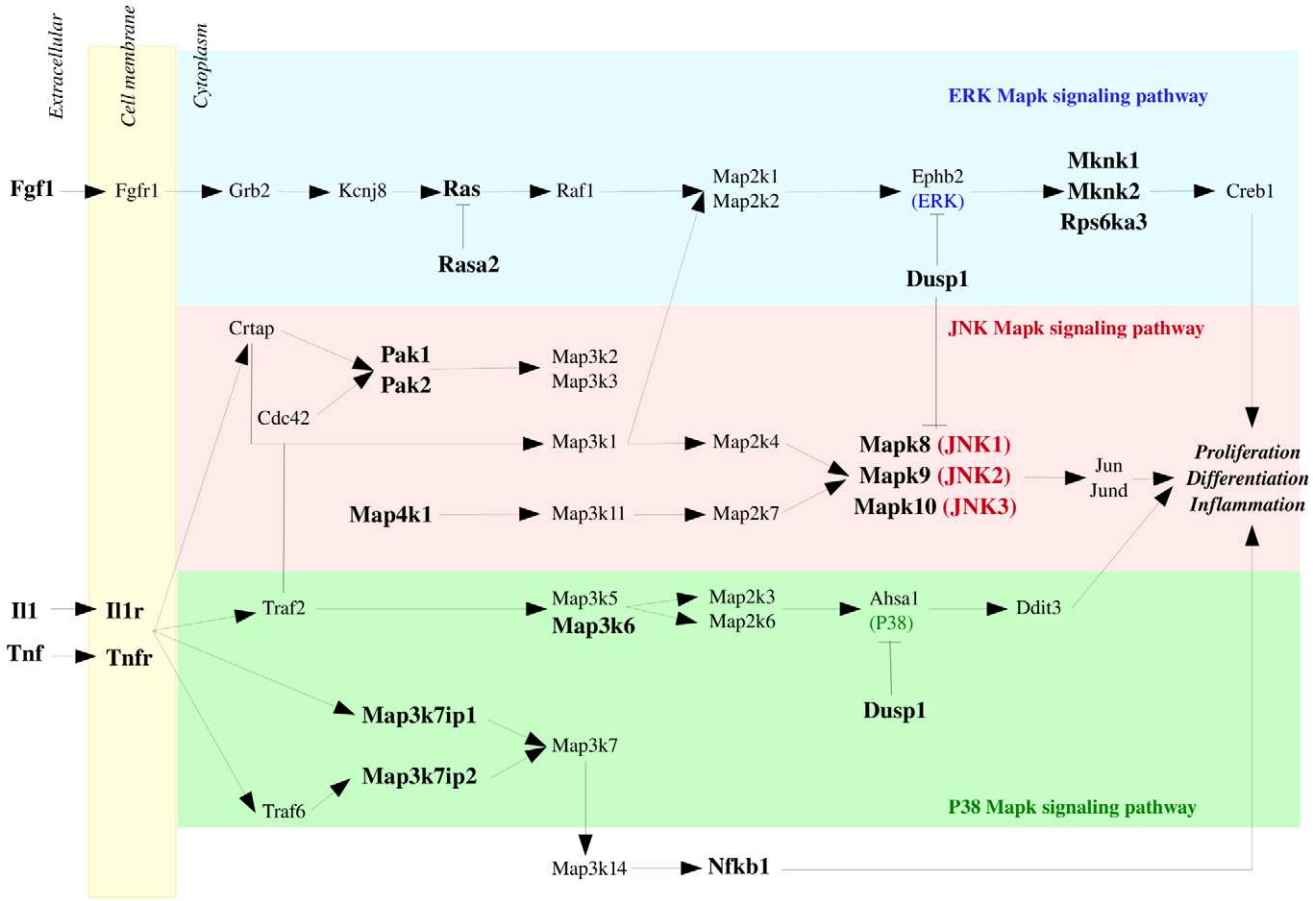


Figure 6. MAPKinase Signaling Pathway. The 3 major groupings of mitogen-activated protein kinase (MAPkinase) pathways are activated rapidly after OA injection. Genes belonging to the extracellular signal-related kinases (ERK) (blue), the Jun amino-terminal kinases (pink) and the p38 proteins (green) are up-regulated early after OA injection. These genes are represented in bold.
doi:10.1371/journal.pone.0011485.g006

esis. Response of cells to Il7 is dependent on the presence of the Il7r [45]. Csf1 induces cells of the monocyte/macrophage lineage [46].

We were able to visualize the spreading of inflammation as many genes involved in leukocyte transendothelial migration and CAM clustered together. Spreading of inflammation is part of the pro-inflammatory response to inflammation. We observed that high expression level of genes involved in the spreading of inflammation signal follows the high expression level of genes involved in inflammation and immune processes. This is probably due to the fact that vascular endothelium is the target of pro-inflammatory cytokines. Once the vascular endothelium activated, genes involved in cell adhesion are expressed and leukocytes are transferred from circulating blood to the site of inflammation.

A late disruption of lipid metabolism

In our study, we found a late disruption of gene expression level related to lipid metabolism. Indeed, among numerous genes whose expression level increases at 3 to 4 hours and reach a peak at 18 hours–24 hours, 21 belong to the lipid metabolic process (figure 5 and see supplementary data file S4). Previous studies showed lipid metabolism disturbance soon after inflammation started. Hardardottir *et al.* have proposed that the changes in lipid and lipoprotein metabolism that occur during the host response to infection/inflammation include anti-infective and anti-inflammatory effects that contribute to the host defense [47]. Infection and inflammation are accompanied by cytokine-induced alterations in lipid and lipoprotein metabolism [48]. As expression level of numerous genes involved in lipid metabolism increased 3 to 4 hours after OA injection and are highly over-expressed 24 hours after OA injection, we believe that lipid metabolism disruption is related to the highly transcribed pro-inflammatory cytokines such as Tnfα and Il1. Indeed, we observed that expression level of Tnfα is maximum 1.5 hours after OA administration (figure 1 B) and Il1, which belongs to cluster Early, is highly expressed 1 hour–1.5 hours after OA administration (figure 4). Il6 is highly expressed 3 hours after OA injection. Moreover, it has been previously shown that serum triglyceride levels, serum total cholesterol and LDL levels increased as a consequence of the highly transcribed pro-inflammatory cytokines [47,49]. Simultaneously, during inflammation, there is a marked decrease in serum levels of HDL [50]. We do not believe that high expression level of genes involved in lipid metabolism is a direct consequence of OA itself. Indeed, even if OA belongs to the fatty acid family, several studies have shown that late disruption of the lipid metabolism is associated with inflammation. Worgall, for example, showed that expression of defective cystic fibrosis transmembrane conductance regulator (CFTR), the cause for cystic fibrosis, is associated with an inflammatory state and affects fatty acid and cholesterol metabolism which is associated, in great majority of patients, with decreased linoleic acid level and increased myristic, palmitoleic, stearic and oleic fatty acids levels [51]. Recently, there has been

much interest in using Statin for sepsis [52]. Statin, which has lipid-lowering properties, reduce total cholesterol, LDL and triglyceride levels [53]. These previous works showed that triggers of inflammation which are not fatty acids led to disruption of lipid metabolism. Feingold *et al.* observed that the hypertriglyceridemic effect of LPS and cytokines is rapid, occurring 2 hours after administration and is sustained for at least 24 hours [16]. Channaiyan *et al.* showed, in a transcriptional study of systemic inflammation in a cecal ligation/puncture model (CLP) of sepsis in rat, that several genes involved in lipid metabolism are up-regulated in lung 6H after CLP [35]. In our study, the timing of the cytokine effect on genes related to lipid metabolism is similar. We showed here that lipid metabolism disturbance is a consequence of inflammation.

Most of the changes in genes related to lipid metabolism that are induced by inflammation are attributable to changes in gene transcription. The expression level of 189 genes involved in at least one out of the 25 GO categories belonging to the transcriptional process are part of the annotation cluster 1 identified by DAVID. They slowly increased 3 hours to 4 hours after OA injection to reach a maximum expression level after 18 hours to 24 hours OA incubation (supplementary table S1). This is probably the consequence of the high expression level of many pro-inflammatory cytokines identified in cluster Early (figure 5). In the promoter region of Il6 and Il1, transcription factors Nfkb1 and Cebpb (NF-IL6) are involved in the expression of many inducible cellular genes that encode cytokines, immunoregulatory receptors, and acute phase proteins [50]. Because many genes involved in regulating immune response and acute phase reaction contain both Cebpb and Nfkb1 sites, it is highly possible that cooperative interactions between Cebpb and Nfkb1 play an important role in the expression of cytokines. For many positive acute phase proteins, transcription activities of their corresponding genes increase during inflammation, reaching a maximum level between 18 hours and 36 hours after inducing an acute inflammation [54]. Our results confirm these findings. Indeed, pro- and anti-inflammatory cytokines highly expressed respectively at early and late stages of the inflammation process induce the transcription of many other genes later after OA injection. Also, Calvano *et al.* showed in their microarray study in 2005 that, in human blood leukocytes, 4 hours to 6 hours after endotoxin injection, the expression of numerous transcription factors was increased. Among them, the expression of several members of the nuclear factor kappa/relA family of transcription factors (Nfkb1, Nfkb2, Rela, and Relb) reached their zenith [34].

We expected a massive release of inflammatory mediators soon after OA injection followed by the release of anti-inflammatory mediators [55]. This sequence of events has been previously described by authors studying sepsis [56]. The response to inflammation we observed and which lasted over 24 hours has been previously described associated with infection [47]. Observing these responses associated with only inflammation justifies the need to fully describe gene expression profiles during inflammation and therefore, it will become possible for other research groups to identify specific responses to infections during sepsis.

Potential targets for therapy

By looking at how changes in gene expression occur through time during lung inflammation, we could confirm or invalidate targets previously identified for therapy.

Among these candidates, Smad7 has been identified [57]. Transient gene transfer and expression of Smad7, introduced by recombinant human type 5 adenovirus vector into the lungs, prevented pulmonary fibrosis induced by bleomycin in mice [57]. Smad7 is an intracellular antagonist of Tgfb signaling [57]. It inhibits Tgfb-induced transcriptional response. Moreover, it is thought that prolonged overproduction of Tgfb induced by repeated chemical or biological injury leads to the accumulation of pathological amounts of extra-cellular matrix in the lung tissue, which is followed by functional deterioration [17]. We believe Smad7 is a good candidate for gene therapy because it belongs to the set of significantly differentially expressed genes across time and its expression level increased 3 to 4 hours after OA injection to reach a maximum expression level 24 hours after injection.

We also considered the angiotensin-converting enzyme gene (Ace) which can predict susceptibility and outcome in ARDS. Circulating Ace derives largely from the pulmonary endothelium. Its release in circulation is affected by extensive endothelial damage [18]. Septic ARDS patients have markedly decreased serum Ace levels compared to those of non-septic ARDS patients [19]. In our non-septic model of ALI, Ace expression level did not significantly change over time.

Finally, we considered transcription of surfactant lipid and surfactant-associated proteins. During ALI in mice, epithelial cell injury leads to reduced surfactant biosynthesis. It has been previously shown that maintenance of surfactant-associated protein B (Sftpb) transcript levels is necessary for over-expression of mouse transforming growth factor α (Tgfa) involved in protection against nikel-induced lung injury [20]. Out of the 10 candidate genes previously identified, 4 are present on MUSV29K: Slc34a, Fasn, Sftpd and Napsa. In our study, Slc34a, involved in surfactant lipid biosynthesis, is the only gene whose expression level changed significantly over time. It belongs to the cluster Late and its expression level increased over time after OA injection to reach a peak at 24 hours incubation.

Our wide-scale transcriptional study identified pathways and specific genes involved in response to inflammation. We found several potential candidates for gene therapy such as Smad7. Since there has been great progress in delivering genes to the lungs over the past decade, gene therapy is expected to become a new tool in addition to classical approaches for treating inherited and acquired diseases including ARDS in the near future.

Conclusion

In this study, we pictured gene expression profiles during inflammation according to the time-course. Overall, our micro-array analysis gives a global overview of critical events occurring during lung inflammation. We have first confirmed that gene-expression profiling discriminates between pro- and anti-inflammatory phases during lung inflammation. The analysis of gene functional annotation reveals several major features. First, it visualizes at transcription level the pro-inflammatory stage characterized by the release of immune response and activation of inflammatory mechanisms. We observed that these pathways are activated soon after the initiation of lung injury. Then the Intermediate cluster revealed that the immune cells migrate into inflamed tissues through interaction with vascular endothelial cells. Finally, at late stage of the inflammation, metabolism is deeply disturbed. We also showed here that inflammation is associated with marked changes in lipid and lipoprotein metabolism suggesting that lipoproteins seem to participate in innate immunity.

This inflammation model in mouse is the first step toward the identification of the molecular mechanisms involved in the infection process at different steps of inflammation. From now on, when studying the impact of an infection at the transcriptional level, we will be able to distinguish between the gene expression level variations due to the pathogen itself and the ones resulting

specifically from the inflammation associated with infection. In our study, by identifying numerous common pathways, we showed that the inflammation alone which is a component of infection has to be considered separately from infection.

Furthermore, it is reasonable to think that the impact of a pathogen on an inflamed organ depends on the pre-existing inflammation. Inoculation of an infectious agent has to take into account timing of the pre-existing inflammation (pro- or anti-inflammatory stage). We reached our first objective of characterizing inflammation across a time-course. Next, we will induce an infection in mouse undergoing inflammation at different time-points (pro- vs. anti- inflammatory phases) using group B streptococcus.

Materials and Methods

Experimental groups

Wild-type female C57Bl/6J mice, 7 weeks old, were obtained from Charles River and housed in a specific pathogen-free animal facility. We studied 35 mice divided into 6 groups (figure 7). Each group included one or two mice administered with physiological serum and 3 to 7 mice administered with OA, which correspond to biological replicates. Each group was identified according to the incubation time: 1 hour, 1.5 hours, 3 hours, 4 hours, 18 hours and 24 hours. All experimental procedures involving animals were approved by the veterinary office of the Ministry of Agriculture, France (authorization number: 13–27).

Injections

Female C57Bl/6J mice were anesthetized with 5% Xylazine–20% Ketamine (0,1 ml/10 g). We injected 20 µl of physiological serum in the control mouse at each time-point and 20 µl of OA (1,2 µl/g body weight, sigma #27728-1L-R) in the other 27 mice of our study. Administration of physiological serum and OA was done through the tail vein with a 0.3-ml insulin syringe (BD #320837).

Sacrifice and organ removal

Each group of mice was sacrificed 1 hour, 1.5 hours, 3 hours, 4 hours, 18 hours or 24 hours after injection of OA or physiological serum by exanguination through the eye vein. For each mouse, we collected the smallest lung for RNA extraction in 4 ml Trizol. The lung was immediately fractionated using a polytron (T8 Ultra-Turrax, IKA WERKE #3420000) and kept at −20°C.

Lung histology

The lung was fixed in 4% Paraformaldehyde and paraffin embedded. Five µm sections were cut and stained with Harri's hematoxylin and eosin (H&E) for morphological examination. Lung inflammation was assessed by studying the following morphological criteria: alveolar/interstitial edema, presence of fibrin, alveolar/interstitial neutrophils. All slides were coded and evaluated in a blinded fashion to prevent bias.

RNA extraction

Total RNA from lung was extracted using TRIzol reagent (Invitrogen, #15596-026). The quality of RNA was confirmed on a 1% denaturing agarose gel, and the concentration of RNA was determined by reading absorbance at 260/280 nm.

Real-Time PCR

Real-time RT-PCR amplification was performed using primer sets for one control gene: Actb (5′-ACTCTTCCAGCCTTCC-TTC-3′ and 5′-ATCTCCTTCTGCATCCTGTC-3′), two pro-inflammatory genes: Tnfα (5′-CCCTCACACTCAGATCATCT-TCT-3′ and 5′-GCTACGACGTGGGCTACAG-3′) and Il6 (5′-TAGTCCTTCCTACCCAATTTCC-3′ and 5′-TTGGTCCTT-AGCCACTCCTTC-3′), and two anti-inflammatory genes: Il10 (5′-GCTCTTACTGACTGGCATGAG-3′ and 5′-CGCAGCT-CTAGGAGCATGTG-3′) and Il4 (5′-TGGATCTGGGAG-CATCAAGGT-3′ and 5′-TGGAAGTGCGGATGTAGTCAG-3′). cDNA was synthesized using 5 µg of RNA, 1 mM dNTP, and1 µg dt25 primer. After 5minutes at 65°C, 10 mM DTT, 40 U RNAaseIN (Promega,#N261B) and 400U of Superscript II transcriptase (Invitrogen, #18064-014) were added. This mix has been maintained 2 hours at 42°C and 15 minutes at 70°C. 1 µl of RNAaseH (Invitrogen,#18021-014) has then been added and the samples have been kept at 37°C for 20 minutes. 1 µl of the RT product was then amplified for 2minutes at 50°C, 10minutes at 95°C followed by 40 cycles (15 sec at 95°C, 30 sec at 60°C, 30 sec at 72°C) followed by one cycle of 72°C for 10minutes and 95°C for 15 sec. This was done in a 25 µl reaction mix containing

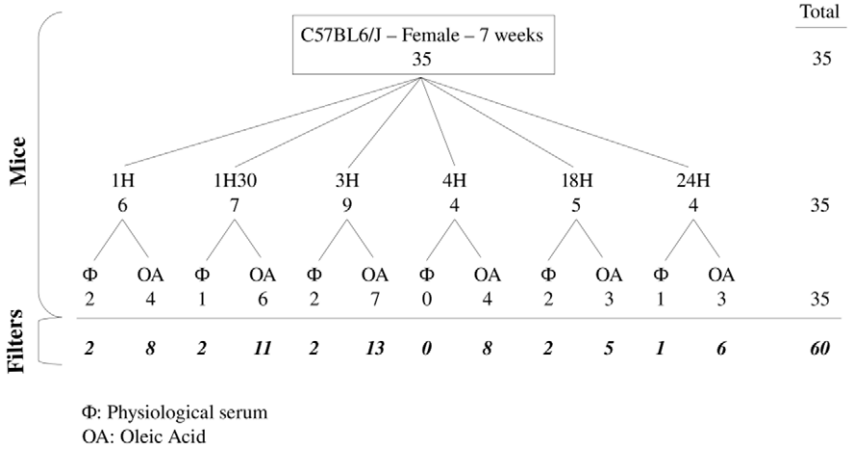


Figure 7. Experimental design of the experiment. We studied 35 mice C57BL6/J, female, 7 weeks old. We divided these mice into 6 groups based on the incubation time (1H, 1H30, 3H, 4H, 18H, and 24H). In each group, mice are divided into 2 subgroups depending on the injection: OA or physiological serum. 25 samples have been hybridized on 2 microarrays consisting in technical replicates and 10 samples have been hybridized once. We performed 60 hybridizations.
doi:10.1371/journal.pone.0011485.g007

1X Power SYBR Green PCR Master Mix (Applied Biosystem, #4367659) and 25pM of each primer using the ABI PRISM 7000 sequence Detection System (Applied Biosystem).

Relative expression of the RT-PCR products was determined using the comparative CT method based on the mathematical model from M.W. Pfaffl [58]. Relative expression values for each pro-inflammatory and anti-inflammatory gene were expressed as a ratio of pro/anti-inflammatory gene expression level after OA injection to the same pro/anti-inflammatory gene expression level after physiological serum at the same time-point. Each sample was run in triplicate. At least 2 negative controls without template or with lung RNA were systematically conducted with each amplification.

Microarray design

cDNA microarrays were designed and prepared as described by Puthier *et al.* [59]. The microarrays used in this study, MUSV29K, contain 7771 sequences. The following cDNA libraries were used: the NIA Mouse 15K cDNA clone set, 2NbMT (thymus), NbMLN (lymph node), and 3NbMS (spleen). Detailed descriptions of these cDNA libraries are available at the UniGene database website (2NbMT: Lib.544, 3NbMS: Lib. 553, NbMLN: Lib.567, NIA 15K: Lib. 8622) [60]. All of the libraries were cloned into pT3T7D-Pac vector, except for the NIA 15K Mouse cDNA clone set, which was cloned into pSPORT1 vector. The NIA 15K Mouse cDNA clone set is a re-arrayed and resequenced set of 15 000 bacterial clones derived from 11 embryo cDNA libraries [61]. 2NbMT, NbMLN and 3NbMS are sequenced I.M.A.G.E libraries. These libraries contain 84127 clones. We used the Microarrays Quality Control tool to select clones matching a single region on the mouse genome and for which at least one 3′ EST and one 5′EST have been identified [62]. We selected 7771 bacterial clones matching 6622 mouse genes. 73% (4833) of the genes are represented by a single cDNA clone and about 27% (1789) of the genes included in this gene set are represented by two or more different cDNA clones, providing internal controls to assess the reproducibility of gene expression measurements. We also added 8 positive controls (poly-A LBP2S and Cot1, a mix of DNA fragments containing repeated sequences) and 2205 negative controls (empty spots and CG03, an *Arabidopsis thaliana* cDNA sequence). In summary, the MUSV29K mouse microarray we designed contains 9984 spots, identifying 6622 mouse genes.

Microarray preparation

PCR amplifications were performed in 96-well microliters plates using the following primers: 5′-CCAGTCACGACGTTG-TAAAACGAC-3′ and 5′-GTGTGGAATTGTGAGCGGATA-ACAA-3′. The reactions were performed as described in Diehl *et al.* [63] by transferring few *Escherichia coli* from a growth culture with a plastic 96pins replicator (Proteigene, #X5055) to 100 µl of PCR mix pH 8,5 containing 1,5 mM MgCl2, 1 M Betain, 375 µM dATP, dTTP, dGTP, dCTP, and 5 U of GoTaq DNA polymerase (Promega, #M3171). The plates were incubated for 6 min at 94°C, before 38 cycles of 94°C for 30 seconds, 65°C for 45 seconds and 72°C for 3.5 minutes, followed by a final elongation phase at 72°C for 10 minutes. Amplification products were not quantified, but their quality was systematically checked on 1% agarose gels. 11,03% of the bacterial clones were estimated to be non-amplified and 4,38% showed multiple bands after PCR amplification. Unpurified PCR products were then transferred to 384-well microplates using the TECAN Genesis workstation 150, before being evaporated. Each well was filled with 30 µl of distilled water, and its content spotted onto nylon membranes (Hybond-N+; Amersham Bioscience). This step was conducted using a MicroGrid-II arrayer (Apogent Discoveries) equipped with a 64-pin biorobotics printhead. Microarrays were hybridized with χdATP 33P-labelled probe, whose oligonucleotide sequence is common to all spotted PCR product (LBP2S 5′-TCACACAG-GAAACAGCTATGAC-3′). It showed uniform signal intensities across individual membranes [64].

Microarray data acquisition and analysis

10 mRNA samples extracted from 10 mice were run on a single microarray. In addition, 25 samples were run on two microarrays, and were considered as technical replicates. This corresponds to a total of 60 microarrays. All microarray procedures were done at our microarray core facility [65]. CDNA were designed and prepared as described in Puthier *et al.* [59], using 5 µg of total RNA in the presence of α-dCTP 33P.

After image acquisition, data were processed as described in figure 3. Hybridization signals were quantified using the Bzscan2 software [66]. All images were carefully inspected and spots with overestimated intensities due to neighborhood effects were manually excluded. We used the NylonArray library for R locally developed by A. Bergon and D. Puthier [67]. This package contains functions to perform diagnosis and normalization of nylon microarray data. For each microarray, we obtained two datasets: the first was obtained by hybridizing the microarray with a probe, whose oligonucleotide sequence is common to all spotted PCR product (vector) and the second was obtained by hybridizing the microarray with lung cDNA (sample). The sample datasets were corrected for neighborhood effects and local background as described by F. Lopez *et al.* [66]. Quantile normalization was applied to sample data to correct for global intensity and dispersion.

Microarray data were statistically analyzed using the TIGR MeV (MultiExperiment Viewer) V4.3.01 software [68]. SAM (Significant Analysis of Microarrays) procedure was applied to look for time specific variation in gene expression in the dataset [14]. A each time-point, mean value of gene expressions in control samples (physiological serum injection) were subtracted from gene expression in OA samples. Hierarchical clustering (average linkage clustering metrics and Pearson correlation for the distance) was applied to the dataset (samples and genes) using the Cluster software and results were visualized with the Treeview software [15]. We identified biological annotation for the clusters using DAVID [69] and interaction networks using the Kegg pathways [70]. We applied a Bonferroni correction to account for multiple tests performed. To interpret our data, we used two functionalities in DAVID: the «Functional annotation» tool and the «Functional annotation clustering» tool. The first one associated gene ID with a biological term which belongs to one out of the 40 annotation categories available in DAVID, including GO terms, protein-protein interactions, protein functional domains, diseases associations and sequence feature. This extended annotation coverage increases the analytic power by allowing investigators to analyze their genes from many different biological aspects in a single space. DAVID functional annotation clustering measured relationships among the annotation terms based on the share of common genes. This type of grouping of functional annotations was able to give a more insightful view of the relationships between annotation categories and terms compared with the traditional linear list of enriched terms, as highly related/redundant annotation terms may be dispersed among hundreds of other terms. Each cluster of functional annotation was associated with an enrichment score, which depended on the distribution of the enrichment (p-values) of each member. A good enrichment score was obtained when most of members had good p-values. This score is a relative score instead of a statistical probability with a minimum and a

maximum value. This means that enrichment scores could be considered only together, by comparing them.

All data are MIAME compliant and have been submitted to the GEO database (http://www.ncbi.nlm.nih.gov/geo/; accession number GSE18712).

Supporting Information

Data file S1 Comparison of FDR5% and FDR0% to identify significantly differentially expressed genes. In our study, we applied a false discovery rate of 5%, which yielded to a set of 1000 genes significantly differentially expressed across the time-course of our experiment. By filtering our data based on $q<0.01$ and FDR 5%, we end up with 121 genes significantly differentially expressed in our experiment. By applying a false discovery rate of 0%, we identify 121 genes significantly differentially expressed across the time-course and identical to the ones identified with FDR 5% and $q<0.01$.
Found at: doi:10.1371/journal.pone.0011485.s001 (0.14 MB PDF)

Data file S2 One thousand genes have been identified by SAM with FDR 5%. Each gene is associated with its expression values in OA samples. Each expression value has been normalized by quantile and adjusted with its physiological serum value at the same time point.
Found at: doi:10.1371/journal.pone.0011485.s002 (0.40 MB TXT)

Data file S3 Identification of clusters of samples using hierarchical clustering. Prior to partition samples (= microarrays) into 3 classes (early, intermediate and late), we performed a hierarchical clustering of the samples in TMEV V4.3.01. using average linkage clustering metrics and Pearson correlation for the distance. This analysis partitioned the samples into 3 groups: early (1H-1H30), intermediate (3H-4H) and late (18H-24H).
Found at: doi:10.1371/journal.pone.0011485.s003 (0.14 MB PDF)

Data file S4 Sample of interesting genes differentially expressed across the time-course. For each cluster, a set of genes mentioned in the manuscript is listed associated with their average expression level at each time-point.
Found at: doi:10.1371/journal.pone.0011485.s004 (0.00 MB TXT)

Table S1 Functional annotation and enrichment of the 366 genes belonging to cluster Late. We identified biological annotation for the cluster Late using DAVID [69] and interaction networks using the Kegg pathways [70]. We applied a Bonferroni correction to account for multiple tests performed. The functional annotation functionality associated gene ID with a biological term which belongs to one out of the 40 annotation categories available in DAVID. DAVID functional annotation clustering measured relationships among the annotation terms based on the share of common genes.
Found at: doi:10.1371/journal.pone.0011485.s005 (0.05 MB DOC)

Acknowledgments

We thank Valérie Gall and Séverine Garnier for technical support. We also thank Pascal Rihet for helpful discussions and Aurélie Bergon for informatics support.

Author Contributions

Conceived and designed the experiments: IL CN. Performed the experiments: IL BL CC SG. Analyzed the data: IL JT. Wrote the paper: IL CN. Advised for data analysis and interpretation: ML.

References

1. Matthay MA, Zimmerman GA, Esmon C, Bhattacharya J, Coller B, et al. (2003) Future research directions in acute lung injury: summary of a National Heart, Lung, and Blood Institute working group. Am J Respir Crit Care Med 167: 1027–1035.
2. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, et al. (2005) A network-based analysis of systemic inflammation in humans. Nature 437: 1032–1037.
3. Bone RC, Grodzin CJ, Balk RA (1997) Sepsis: a new hypothesis for pathogenesis of the disease process. Chest 112: 235–243.
4. Gogos CA, Drosou E, Bassaris HP, Skoutelis A (2000) Pro- versus anti-inflammatory cytokine profile in patients with severe sepsis: a marker for prognosis and future therapeutic options. J Infect Dis 181: 176–180.
5. Cavaillon J (1998) Pathophysiological rôle of pro- and anti-inflammatory cytokines in sepsis. Dans: The diagnosis of sepsis; current prospectives. Boston: Kluwer Academic Publishers. pp 127–140.
6. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, et al. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531–537.
7. Wurfel MM (2007) Microarray-based analysis of ventilator-induced lung injury. Proc Am Thorac Soc 4: 77–84.
8. Saban MR, Hellmich H, Nguyen NB, Winston J, Hammond TG, et al. (2001) Time course of LPS-induced gene expression in a mouse model of genitourinary inflammation. Physiol Genomics 5: 147–160.
9. Wang Z, Beach D, Su L, Zhai R, Christiani DC (2008) A genome-wide expression analysis in blood identifies pre-elafin as a biomarker in ARDS. Am J Respir Cell Mol Biol 38: 724–732.
10. Schuster DP (1994) ARDS: clinical lessons from the oleic acid model of acute lung injury. Am J Respir Crit Care Med 149: 245–260.
11. Derks CM, Jacobovitz-Derks D (1977) Embolic pneumopathy induced by oleic acid. A systematic morphologic study. Am J Pathol 87: 143–158.
12. Zhou Z, Kozlowski J, Schuster DP (2005) Physiologic, biochemical, and imaging characterization of acute lung injury in mice. Am J Respir Crit Care Med 172: 344–351.
13. Ulrich K, Stern M, Goddard ME, Williams J, Zhu J, et al. (2005) Keratinocyte growth factor therapy in murine oleic acid-induced acute lung injury. Am J Physiol Lung Cell Mol Physiol 288: L1179–1192.
14. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98: 5116–5121.
15. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863–14868.
16. Feingold KR, Staprans I, Memon RA, Moser AH, Shigenaga JK, et al. (1992) Endotoxin rapidly induces changes in lipid metabolism that produce hypertriglyceridemia: low doses stimulate hepatic triglyceride production while high doses inhibit clearance. J Lipid Res 33: 1765–1776.
17. Border WA, Noble NA (1994) Transforming growth factor beta in tissue fibrosis. N Engl J Med 331: 1286–1292.
18. Villar J, Flores C, Pérez-Méndez L, Maca-Meyer N, Espinosa E, et al. (2008) Angiotensin-converting enzyme insertion/deletion polymorphism is not associated with susceptibility and outcome in sepsis and acute respiratory distress syndrome. Intensive Care Med 34: 488–495.
19. Casey L, Krieger B, Kohler J, Rice C, Oparil S, et al. (1981) Decreased serum angiotensin converting enzyme in adult respiratory distress syndrome associated with sepsis: a preliminary report. Crit Care Med 9: 651–654.
20. Bein K, Wesselkamper SC, Liu X, Dietsch M, Majumder N, et al. (2009) Surfactant- associated protein B is critical to survival in nickel-induced injury in mice. Am J Respir Cell Mol Biol 41: 226–236.
21. Bem RA, van Woensel JBM, Bos AP, Koski A, Farnand AW, et al. (2009) Mechanical ventilation enhances lung inflammation and caspase activity in a model of mouse pneumovirus infection. Am J Physiol Lung Cell Mol Physiol 296: L46–56.
22. Cavaillon J, Adib-Conquy M, Fitting C, Adrie C, Payen D (2003) Cytokine cascade in sepsis. Scand J Infect Dis 35: 535–544.
23. Cohen J (2002) The immunopathogenesis of sepsis. Nature 420: 885–891.
24. Bhatia M, Moochhala S (2004) Role of inflammatory mediators in the pathophysiology of acute respiratory distress syndrome. J Pathol 202: 145–156.
25. Nijsten MW, Hack CE, Helle M, ten Duis HJ, Klasen HJ, et al. (1991) Interleukin-6 and its relation to the humoral immune response and clinical parameters in burned patients. Surgery 109: 761–767.
26. Mizuhara H, O'Neill E, Seki N, Ogawa T, Kusunoki C, et al. (1994) T cell activation-associated hepatic injury: mediation by tumor necrosis factors and protection by interleukin 6. J Exp Med 179: 1529–1537.

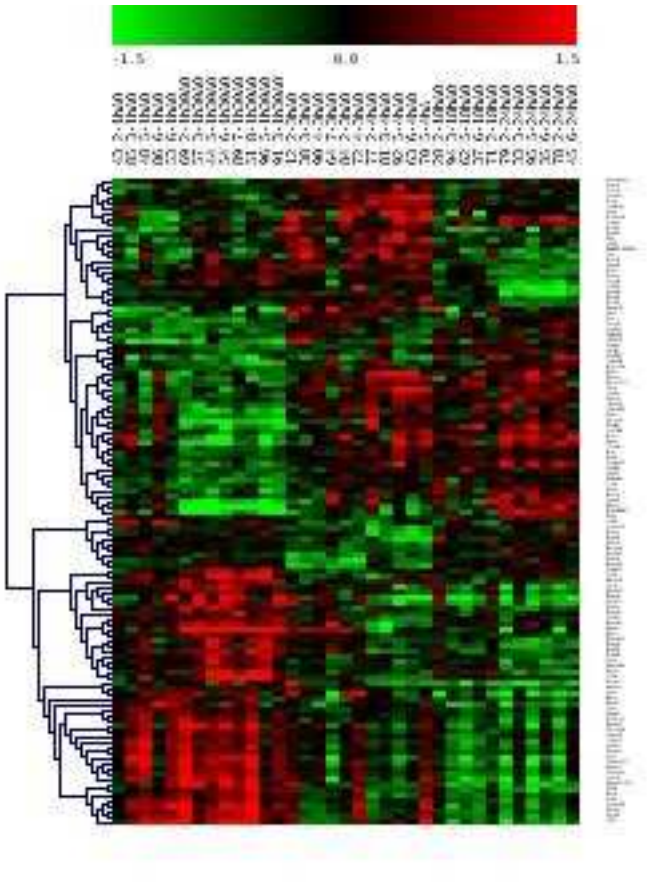
27. Crowl RM, Stoller TJ, Conroy RR, Stoner CR (1991) Induction of phospholipase A2 gene expression in human hepatoma cells by mediators of the acute phase response. J Biol Chem 266: 2647–2651.
28. Akira S, Taga T, Kishimoto T (1993) Interleukin-6 in biology and medicine. Adv Immunol 54: 1–78.
29. Kasama T, Strieter RM, Lukacs NW, Burdick MD, Kunkel SL (1994) Regulation of neutrophil-derived chemokine expression by IL-10. J Immunol 152: 3559–3569.
30. Howard M, O'Garra A, Ishida H, de Waal Malefyt R, de Vries J (1992) Biological properties of interleukin 10. J Clin Immunol 12: 239–247.
31. Armstrong L, Millar AB (1997) Relative production of tumour necrosis factor alpha and interleukin 10 in adult respiratory distress syndrome. Thorax 52: 442–446.
32. te Velde AA, Huijbens RJ, Heije K, de Vries JE, Figdor CG (1990) Interleukin-4 (IL-4) inhibits secretion of IL-1 beta, tumor necrosis factor alpha, and IL-6 by human monocytes. Blood 76: 1392–1397.
33. Rossi D, Zlotnik A (2000) The biology of chemokines and their receptors. Annu Rev Immunol 18: 217–242.
34. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, et al. (2005) A network-based analysis of systemic inflammation in humans. Nature 437: 1032–1037.
35. Chinnaiyan AM, Huber-Lang M, Kumar-Sinha C, Barrette TR, Shankar-Sinha S, et al. (2001) Molecular Signatures of Sepsis: Multiorgan Gene Expression Profiles of Systemic Inflammation.
36. Olman MA, White KE, Ware LB, Simmons WL, Benveniste EN, et al. (2004) Pulmonary edema fluid from patients with early lung injury stimulates fibroblast proliferation through IL-1 beta-induced IL-6 expression. J Immunol 172: 2668–2677.
37. Bulfone-PauS S, Bulanova E, Pohl T, Budagian V, Durkop H, et al. (1999) Death deflected: IL-15 inhibits TNF-alpha-mediated apoptosis in fibroblasts by TRAF2 recruitment to the IL-15Ralpha chain. FASEB J 13: 1575–1585.
38. Muller WA (2009) Mechanisms of transendothelial migration of leukocytes. Circ Res 105: 223–230.
39. Leone M, Garcin F, Chaabane W, Boutière-Albanèse B, Albanèse J, et al. (2003) Activation of adhesion molecules in patients with septic shock]. Ann Fr Anesth Reanim 22: 721–729.
40. Johnson P, Ruffell B (2009) CD44 and its role in inflammation and inflammatory diseases. Inflamm Allergy Drug Targets 8: 208–220.
41. Woodfin A, Voisin M, Nourshargh S (2007) PECAM-1: a multi-functional molecule in inflammation and vascular biology. Arterioscler Thromb Vasc Biol 27: 2514–2523.
42. Adams DH, Nash GB (1996) Disturbance of leucocyte circulation and adhesion to the endothelium as factors in circulatory pathology. Br J Anaesth 77: 17–31.
43. Giannoudis PV, Smith RM, Banks RE, Windsor AC, Dickson RA, et al. (1998) Stimulation of inflammatory markers after blunt trauma. Br J Surg 85: 986–990.
44. Opdenakker G (2001) New insights in the regulation of leukocytosis and the role played by leukocytes in septic shock. Verh K Acad Geneeskd Belg 63: 531-538; discussion 538-541.
45. Huang H, Chiang Y, Hung S, Li C, Yen JJ (2007) An IL-7 splicing-defect lymphopenia mouse model revealed by genome-wide mutagenesis. J Biomed Sci 14: 169–181.
46. Loureiro RMB, Monaco K, Kearney JB, Blickarz-Durand CE, Kirby SL, et al. (2008) csf1 is required for early embryonic macrophage development: characterization of the csf1(op)/csf1(op) mutation in ES cell-derived macrophages. Br J Haematol 141: 739–742.
47. Hardardóttir I, Grunfeld C, Feingold KR (1995) Effects of endotoxin on lipid metabolism. Biochem Soc Trans 23: 1013–1018.
48. Argilés JM, Lopez-Soriano FJ, Evans RD, Williamson DH (1989) Interleukin-1 and lipid metabolism in the rat. Biochem J 259: 673–678.
49. Grunfeld C, Soued M, Adi S, Moser AH, Fiers W, et al. (1991) Interleukin 4 inhibits stimulation of hepatic lipogenesis by tumor necrosis factor, interleukin 1, and interleukin 6 but not by interferon-alpha. Cancer Res 51: 2803–2807.
50. Baumann H, Prowse KR, Marinković S, Won KA, Jahreis GP (1989) Stimulation of hepatic acute phase response by cytokines and glucocorticoids. Ann N Y Acad Sci 557: 280-295, discussion 295-296.
51. Worgall TS (2009) Lipid metabolism in cystic fibrosis. Curr Opin Clin Nutr Metab Care 12: 105–109.
52. Woerndle RH, Maxwell DL (2008) Statins and sepsis: good bullet, disappearing target. J Am Osteopath Assoc 108: 486–490.
53. Kopterides P, Falagas ME (2009) Statins for sepsis: a critical and updated review. Clin Microbiol Infect 15: 325–334.
54. Birch HE, Schreiber G (1986) Transcriptional regulation of plasma protein synthesis during inflammation. J Biol Chem 261: 8077–8080.
55. Hotchkiss RS, Coopersmith CM, McDunn JE, Ferguson TA (2009) The sepsis seesaw: tilting toward immunosuppression. Nat Med 15: 496–497.
56. Johnson SB, Lissauer M, Bochicchio GV, Moore R, Cross AS, et al. (2007) Gene expression profiles differentiate between sterile SIRS and early sepsis. Ann Surg 245: 611–621.
57. Nakao A, Afrakhte M, Morén A, Nakayama T, Christian JL, et al. (1997) Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. Nature 389: 631–635.
58. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45.
59. Puthier D, Joly F, Irla M, Saade M, Victorero G, et al. (2004) A general survey of thymocyte differentiation by transcriptional analysis of knockout mouse models. J Immunol 173: 6109–6118.
60. Unigene database website - Library browser: http://www.ncbi.nih.gov/UniGene/lbrowse2.cgi?TAXID_10090. Accessed 13 September.
61. Tanaka TS, Kunath T, Kimber WL, Jaradat SA, Stagg CA, et al. (2002) Gene expression profiling of embryo-derived stem cells reveals candidate genes associated with pluripotency and lineage specificity. Genome Res 12: 1921–1928.
62. Granjeaud S, Lopez F, Hingamp P MiQC: http://tagc.univ-mrs.fr/welcome/spip/.php?rubrique120. Accessed 27 November 2007.
63. Diehl F, Beckmann B, Kellner N, Hauser NC, Diehl S, et al. (2002) Manufacturing DNA microarrays from unpurified PCR products. Nucleic Acids Res 30: e79.
64. Tanaka TS, Jaradat SA, Lim MK, Kargul GJ, Wang X, et al. (2000) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray. Proc Natl Acad Sci U S A 97: 9127–9132.
65. TAGC website: http://www.tagc.univ-mrs.fr/. Accessed 12 September 2007.
66. Lopez F, Rougemont J, Loriod B, Bourgeois A, Loï L, et al. (2004) Feature extraction and signal processing for nylon DNA microarrays. BMC Genomics 5: 38.
67. Puthier D, Bergon A NylonArray: ftp://tagc.univ-mrs.fr/public.bzscan/nylonArray/Accessed 17 April 2008.
68. TMeV: http://www.tm4.org/mev.html. Accessed 7 April 2008.
69. Sherman BT, Huang DW, Tan Q, Guo Y, Bour S, et al. (2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics 8: 426.
70. KEGG: http://www.genome.ad.jp/kegg/pathway.html. Accessed 7 April 2008.

In our study, we applied a false discovery rate of 5%, which yielded to a set of 1000 genes significantly differentially expressed across the time-course of our experiment.
By selecting among these 1000 genes, the ones with q<0.01 and FDR 5%, we ended up with 121 genes significantly differentially expressed in our experiment (see supplementary data file #4).
At the same time, we applied a false discovery rate of 0% and then identified 121 genes significantly differentially expressed across the time-course and identical to the ones identified with FDR 5% and q<0.01.
We performed unsupervised hierarchical clustering on these 121 genes and identified 3 clusters:

−cluster 1: genes over-expressed at 1-1H30

−cluster 2: genes transitionally over-expressed at 3-4H

−cluster 3: genes over-expressed at 18-24H



To analyze functional annotation of genes belonging to each cluster, we used DAVID V6.7. In cluster Early, GO terms related immune response such as to regulation of T cell differentiation (3,2.10-3) and lymphocyte activation (1,3.10-2) were highly represented.
Most of the genes grouped in the cluster Intermediate were found to be involved in regulation of transcription (7,9.10-3).
Finally, functional annotation analysis of genes belonging to the cluster Late showed a slight over-representation of GO terms related to chromatin modification (7,4.10-2).

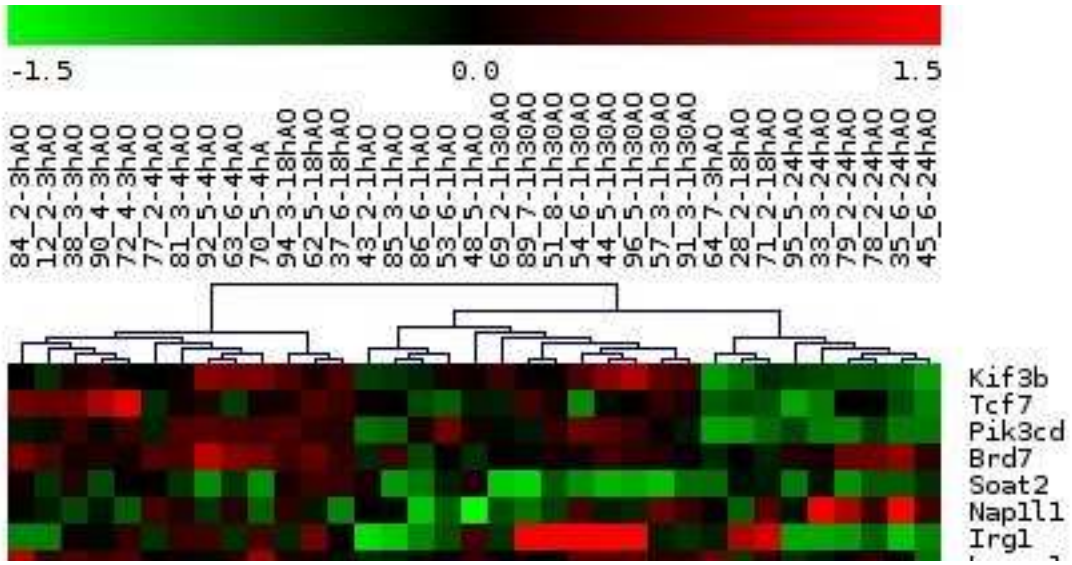
We can conclude that, by limiting our analysis to a very small set of genes (121 genes with FDR5% and q<0.01 = 121 genes with FDR 0%), we lost most of the information initially contained in the 1000 genes dataset filtered with FDR 5%. It has to be noted that, using FDR 5% and q<0.01, the 121 genes are still grouped into 3 clusters. Functional annotations of the 121 genes are very unspecific and could be found in many other studies. Furthermore, several pathways such as modification of lipid metabolism and migration of the inflammatory signal, which are meaningful because supported by previous studies, are swallowed up by very general descriptions.

Data file S2.

One thousand genes have been identified by SAM with FDR 5%. Each gene is associated with its expression values in OA samples. Each expression value has been normalized by quantile and adjusted with its physiological serum value at the same time point.

Ce fichier n'est pas reproduit compte tenu du nombre important de pages qu'il représente (tableau contenant les valeurs d'expression au sein de chaque échantillon des 1000 gènes identifiés).

Prior to partition samples (=microarrays) into 3 classes (early, intermediate and late), we performed a hierarchical clustering of the samples in TMEV V4.3.01. using average linkage clustering metrics and Pearson correlation for the distance.



We can see that samples at 1H and 1H30 clustered together. The same way, samples at 3H and 4H clearly clustered together. Finally, samples at 18H and 24H clustered together with the exception of three 18H samples which have been grouped with the 3-4H cluster as an out-group.

This result support our decision to partition the samples into 3 groups: early (1H-1H30), intermediate (3H-4H) and late (18H-24H).

MGI_ID	Moy_1H	Moy_1H30	Moy_3H	Moy_4H	Moy_18H	Moy_24H
Cluster B – 1H-1H30						
Chemokines						
Cxcl5	-0,60	0,54	-0,05	-0,31	-0,83	-0,60
Ccl22	0,18	0,53	0,17	0,31	-0,04	-0,34
Ccl4	-0,20	1,15	-0,05	-0,35	-0,28	-0,05
Ccl3	0,14	1,63	-0,17	-0,69	0,12	-0,13
Ccl6	-0,48	0,49	-0,25	-0,27	0,48	-1,17
Ccl25	0,21	0,07	-0,12	-0,31	0,25	0,17
MapK signaling pathway						
Fgf1	0,37	0,20	0,44	0,39	-0,17	0,05
Rasa2	-0,24	0,33	0,20	0,21	-0,05	-0,36
Rras	-0,03	-0,20	-0,36	-0,01	-0,13	-0,26
Dusp1	1,27	0,90	0,29	-0,39	0,66	-1,16
Mknk1	-0,30	0,52	-0,16	-0,56	0,03	-0,49
Mknk2	0,18	0,01	-0,86	1,83	0,42	-0,59
Rps6ka3	0,23	0,19	-0,08	-0,27	0,16	-0,15
Mapk8ip3		0,50	-0,05	-0,25	0,34	0,24	0,21
Mapk9	0,12	0,39	0,06	-0,20	-0,09	0,23
Il1b	-0,66	1,50	-0,59	0,10	0,8	-3,16
Map3k7ip1		0,08	0,54	-0,28	0,07	0,09	-0,18
Map3k7ip2		0,16	1,11	0,02	-0,35	-0,64	-0,96
Map3k6	-0,11	0,50	-0,16	-0,52	0,08	0,20	
Il15ra	0,15	0,32	0,11	-0,44	0,03	0,30
Il12rb2	-0,27	0,11	-0,30	0,12	-0,37	-0,74
Il12rb2	-0,18	0,16	-0,03	-0,56	0,04	-0,23
Cluster A – 3-4H						
Transmembrane and endothelial cell adhesion						
Cd44	-0,78	0,73	0,10	1,34	0,31	-1,20
Cd44	-0,83	0,02	0,16	0,01	0,07	0,73
Pecam1	0,61	-0,43	-0,05	0,18	0,14	-0,09
Pecam1	-0,07	-0,38	0,07	1,30	-0,1	-0,05
Icam1	0,18	0,60	-0,01	0,91	-0,03	-0,86
Cd34	-0,56	-0,30	-0,08	0,87	-0,08	0,19
Cd34	0,39	0,37	0,13	-0,62	-0,13	0,30
Glycam1	0,05	-0,09	0,18	0,66	0,27	0,00
Blood cell synthesis						
Cd34	-0,56	-0,30	-0,08	0,87	-0,08	0,19
Cd34	0,39	0,37	0,13	-0,62	-0,13	0,30
Il7r	-0,38	0,38	0,51	0,99	0	-0,18
Il7r	-0,19	0,29	0,46	0,53	0,23	0,00
Csf1	0,15	0,42	-0,17	0,07	-0,16	-0,06
Csf1	0,12	0,56	-0,23	-0,33	-0,14	-0,77
Cd38	-0,70	-0,78	-0,08	1,06	-0,06	-0,09
Cluster C – 18-24H						
Lipid metabolism						
Soat2	-0,36	-0,96	-0,17	-0,45	0,18	-0,51
Hdlbp	0,20	-0,05	-0,21	0,38	-0,09	-0,13
Hdlbp	0,14	-0,31	-0,13	-0,39	0,1	0,29
Hdlbp	0,46	-0,29	-0,25	-0,22	-0,06	0,67
Mbtps1	0,13	-0,31	-0,06	0,07	0,3	0,16
Cyb5r3	-0,11	-0,56	0,11	0,12	0	-0,29
Nsdhl	0,16	-0,20	0,08	-0,26	0,22	0,48
Osbpl5	-0,08	-0,27	-0,19	0,18	0,22	0,35
Il4	-0,75	-0,12	0,32	-0,23	0,01	0,22
Il4ra	-0,21	0,33	0,04	0,26	0,11	-0,04
Hsd3b1	0,19	-0,11	0,16	-0,15	0,45	0,99
Osbpl11	-0,02	-0,35	0,16	0,50	0,06	0,53
Ppara	0,08	-0,79	-0,67	0,41	0,21	0,71
Pigv	0,11	-0,69	-0,02	0,23	0,15	0,36
Ltc4s	0,00	-0,41	-0,32	-0,30	0,34	0,25
D5Wsu178e		-0,26	0,12	-0,21	-1,38	-0,24	-1,02
D5Wsu178e		0,06	-0,23	-0,32	-0,07	0,13	0,29
Ptdss1	-0,34	-0,07	-0,30	-0,18	0,18	0,27
Agps	0,10	-1,34	-0,23	0,81	0,12	0,93
Pisd	-0,58	-0,55	-0,02	0,17	0,06	0,25
A4galt	0,07	-0,81	-0,05	-0,10	0,05	0,34
A4galt	0,54	-1,46	0,33	0,23	0,47	1,05
Acaa2	0,35	-0,64	-0,07	-0,07	0,21	0,72
Acadl	0,14	-0,36	-0,11	-0,28	0,3	-0,14

Acadl	0,08	-0,15	0,02	-0,19	0,32	0,27
Serinc1	-0,18	-0,68	0,41	0,16	-0,49	1,06
Abhd4	0,00	-0,13	-0,24	-0,13	-0,06	0,65
Prei4	0,06	-0,14	-0,15	0,09	0,26	-0,19
Pck2	0,09	-0,09	0,17	-0,12	0,17	0,85
Pfkfb2	0,18	-0,64	-0,01	0,02	0,37	0,72
Th	0,10	-0,08	-0,27	-0,08	0,28	0,50

Table S1. Functional annotation and enrichment of the 366 genes belonging to cluster Late.

Functional classification			
Database	*Term*	*Count*	*P-Value*
Gene ontology	DNA metabolic process	32	7.10E-04
Gene ontology	Regulation of transcription	63	1.10E-03
Gene ontology	Transcription. DNA-dependent	60	1.30E-03
Gene ontology	Regulation of transcription. DNA dependent	59	1.40E-03
Gene ontology	Transcription	65	1.40E-03
Gene ontology	Metabolic process	186	2.20E-03
Gene ontology	Cholesterol metabolic process	7	9.30E-03
Annotation Cluster 1	**Enrichment score: 2.07**		
Database	*Term*	*Count*	*P-Value*
Gene ontology	Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	65	7.70E-04
Gene ontology	Regulation of transcription	63	1.10E-03
Gene ontology	Transcription, DNA-dependent	60	1.30E-03
Gene ontology	Regulation OF transcription, DNA-dependent	59	1.40E-03
Gene ontology	Transcription	65	1.40E-03
Gene ontology	Regulation of gene expression	65	3.20E-03
Gene ontology	Transcription factor activity	26	4.30E-02
Gene ontology	Transcription regulator activity	34	9.00E-02
Annotation Cluster 2	**Enrichment score: 1.96**		
Database	*Term*	*Count*	*P-Value*
Gene ontology	Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	93	1.90E-03
Gene ontology	Metabolic process	186	2.20E-03
Gene ontology	Cellular metabolic process	173	2.30E-03
Gene ontology	Primary metabolic process	172	4.50E-03
Gene ontology	Macromolecule metabolic process	145	5.00E-02
Gene ontology	Cellular process	229	7.10E-02
Annotation Cluster 3	**Enrichment score: 1.44**		
Database	*Term*	*Count*	*P-Value*
Gene ontology	Cholesterol metabolic process	7	9.30E-03
Gene ontology	Sterol metabolic process	7	1.10E-02

Gene ontology	Steroid metabolic process	9	1.80E-02
Gene ontology	Cellular lipid metabolic process	20	5.20E-02
Gene ontology	Alcohol metabolic process	12	6.70E-02
Gene ontology	Lipid metabolic process	21	9.40E-02

An Evaluation of the Role of Gene Expression in the Prediction and Diagnosis of Ventilator-associated Pneumonia

Julien Textoris, Béatrice Loriod, Laurent Benayoun, Pierre-Antoine Gourraud, Denis Puthier, Jacques Albanèse, Jean Mantz, Claude Martin, Catherine Nguyen, Marc Leone.

Anesthesiology, (in press)

Les outils d'analyse haut-débit du transcriptome permettent d'explorer les voies métaboliques mises en jeu en pathologie et d'identifier de potentiels marqueurs diagnostiques. Cette étude a analysé les modifications du transcriptome sanguin lors de la survenue d'une pneumonie associée à la ventilation mécanique (PAVM), chez des sujets admis en réanimation pour traumatisme grave.

Cette étude multicentrique a inclus de manière prospective 165 patients, dont 41 ont développé une PAVM. Des échantillons de sang total ont été prélevés à l'admission, et au moment du diagnostic de PAVM. Afin d'identifier de potentiels marqueurs pronostiques de survenue de PAVM, nous avons comparé par puce à ADN le transcriptome des prélèvements initiaux des patients qui ont développé ou non une PAVM. Afin d'identifier des marqueurs diagnostiques, nous avons comparé le transcriptome des prélèvements appariés des patients qui ont développé une PAVM. Des patients traumatisés qui n'ont pas développé de PAVM ont été utilisés comme contrôles afin d'éliminer un effet du temps. Les résultats ont été confirmés par une autre technologie (NanoString®).

Les patients traumatisés dont l'évolution s'est compliquée d'une PAVM avaient un score lésionnel (ISS pour « Injury Severity Score ») plus élevé (36 *vs.* 29) et un traumatisme thoracique plus fréquent (28 (68 %) *vs.* 57 (46 %)). Aucun marqueur transcriptionnel pronostique n'a été identifié. L'analyse des patients qui présentaient une PAVM a identifié une signature transcriptionnelle entre l'admission et le diagnostic de PAVM. L'analyse de sujets contrôles a montré que cette modulation est liée à l'évolution du traumatisme et non au processus infectieux. Toutefois, une diminution de l'expression de certains médiateurs pro-inflammatoires est observée uniquement chez les patients qui développent une PAVM.

Contrairement à la clinique, l'analyse du transcriptome sanguin ne permet pas de prédire ou de diagnostiquer la survenue d'une PAVM chez des patients traumatisés. Il semble difficile de distinguer, au niveau du transcriptome, inflammation et infection en pratique clinique. Toutefois, il serait intéressant d'évaluer dans un travail complémentaire le rôle de la diminution des médiateurs pro-inflammatoires dans la survenue des PAVM.

An Evaluation of the Role of Gene Expression in the Prediction and Diagnosis of Ventilator-associated Pneumonia

Julien Textoris, M.D.,* Béatrice Loriod, B.S.,† Laurent Benayoun, M.D.,‡ Pierre-Antoine Gourraud, Ph.D.,§ Denis Puthier, M.D.,‖ Jacques Albanèse, M.D., Ph.D.,# Jean Mantz, M.D., Ph.D.,** Claude Martin, M.D.,†† Catherine Nguyen, Ph.D.,‡‡ Marc Leone, M.D., Ph.D.#

ABSTRACT

Background: The Sepsi*Chip* project explored transcriptional modulation associated with ventilator-associated pneumonia (VAP) in patients admitted to the intensive care unit for trauma. Genome-wide expression analysis may help to identify potential diagnostic markers for diseases. The current study examined the changes in blood transcriptome during VAP.

Methods: The authors prospectively included 165 trauma patients, and 41 developed VAP. Whole blood samples were collected at admission and at VAP. To predict VAP, the admission samples were compared by microarray in patients who did or did not develop VAP. To identify diagnosis markers, paired samples of 35 patients who developed VAP were analyzed. Using NanoString (Seattle, WA), the results were confirmed in the patients who developed VAP. Trauma patients who did not develop VAP served as controls to eliminate a time effect.

Results: The injury severity scores of the patients who did or did not develop VAP were 36 and 29, respectively. No predictive biomarker was identified. For patients who developed VAP, a transcriptional signature was identified between the two sampling times. However, this signature was a generalized pattern related to trauma, independent of the infectious process. Genes involved in the proinflammatory response were down-regulated in the patients who developed VAP, but this difference was not statistically significant.

Conclusions: In contrast to clinical assessment, transcriptional analysis of whole blood samples cannot predict or diagnose VAP in trauma patients. Differentiating infection from inflammation seems challenging.

What We Already Know about This Topic

- The clinical diagnostic criteria for ventilator-associated pneumonia (VAP) do not exclude patients who do not have VAP (true negatives)

What This Article Tells Us That Is New

- Transcriptional analysis, using DNA microarrays containing 8,682 genes, of 165 trauma patient's blood was done in patients with (41) and without (42) VAP but could not distinguish inflammation from infection

IN trauma patients, nosocomial infections are a common cause of morbidity and mortality. Pulmonary infection is one of the most life-threatening infections.[1] Prompt initiation of appropriate antimicrobial therapy is critical in the care of

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal's Web site (www.anesthesiology.org).

* Fellow, Service d'Anesthésie et de Réanimation and Centre d'Investigation Clinique, Hôpital Nord, Assistance Publique - Hôpitaux de Marseille, Université de la Méditerranée, Marseille, France; Technologies Avancées pour le Génome et la Clinique (TAGC), unité INSERM U928, Université de la Méditerranée, Marseille. † Research Engineer, ‖ Assistant Professor, ‡‡ Professor and Chairman, Technologies Avancées pour le Génome et la Clinique (TAGC), unité INSERM U928, Université de la Méditerranée. ‡ Assistant Professor, ** Professor and Chairman, Service d'Anesthésie et de Réanimation, Hôpital Beaujon, Assistance Publique - Hôpitaux de Paris, Clichy, France. § Assistant Professor, INSERM U558, Université de Toulouse, Toulouse, France. # Professor, †† Professor and Chairman, Service d'Anesthésie et de Réanimation and Centre d'Investigation Clinique, Hôpital Nord, Assistance Publique - Hôpitaux de Marseille, Université de la Méditerranée.

Received from Service d'Anesthésie et de Réanimation, Hôpital Nord, Assistance Publique - Hôpitaux de Marseille, Université de la Méditerranée, Marseilles, France. Submitted for publication September 22, 2009. Accepted for publication May 5, 2011. Supported by grants from Conseil Régional Provence Alpes Côte d'Azur, Marseilles, France; Société Française d'Anesthésie et de Réanimation, Paris, France; and le Ministère de la Santé (Programme Hospitalier de Recherche Clinique, PHRC 2005), Paris, France. Preliminary results were reported at the 50th annual meeting of the Société Française d'Anesthésie et de Reanimation, Paris, France, September 30, 2008.

Drs. Catherine Nguyen and Marc Leone contributed equally to this work.

Address correspondence to Dr. Leone: Service d'Anesthésie et de Réanimation, Hôpital Nord, Chemin des Bourrelly, 13915 Marseille Cedex 20, France. marc.leone@ap-hm.fr. Anesthesiology 2011; 115:1–1

these patients.[2] Unfortunately, trauma induces a major inflammatory response, making the diagnosis of ventilator-associated pneumonia (VAP) difficult. Chest imaging often is irrelevant because of contusions. In this setting, clinical scores are not informative.[3] Thus, biologic markers may have a role in predicting or diagnosing VAP in patients. The concentrations of procalcitonin can reinforce the diagnosis of infection, but its value as a diagnostic tool remains a matter of debate.[4,5] Several high-throughput studies suggest that genome-wide expression analysis could help with the diagnosis of VAP.[6]

High-throughput technologies may prove to be helpful for deciphering the complex host response to infection. These technologies may provide tools for predicting the occurrence of infection or facilitating its diagnosis. Using genome-wide expression analysis, the signature profiles of inflammation have been characterized by many groups (*e.g.*, trauma,[7–9] lipopolysaccharide infusion,[10] malaria,[11] septic shock in children,[12,13] severe sepsis and septic shock in adults,[6,14–17] and Gram-specific infections[6,18,19]).

To investigate the host response to infection, we attempted to reduce the background noise of inflammation by using trauma patients who developed VAP as their own controls. We prospectively included trauma patients and collected whole blood samples for transcriptome analysis. We conducted an unbiased analysis by exploring the transcriptome using DNA-microarray technology. We compared samples taken at the time of admission from trauma patients who did or did not develop VAP to identify prognostic markers. We also compared paired samples taken from trauma patients at the time of admission and at the time of diagnosis of VAP to identify potential diagnostic markers of VAP. Finally, to eliminate general gene expression patterns related to the elapse of time after trauma, we collected late samples from control trauma patients who did not develop VAP. To confirm our microarray findings, we analyzed the transcriptome with an alternative technology (NanoString nCounter Analysis System; NanoString Technologies, Seattle, WA). We used this technology to confirm our potential gene expression signature and to investigate the expression of some already known inflammatory markers.

Materials and Methods

Patients

The study protocol was approved by the Ethics Committee (Comité de Protection des Personnes Sud-Méditerranée II, no. 206–005). All trauma patients admitted to two surgical intensive care units (ICU) in teaching hospitals (Hôpital de Beaujon, Clichy, France; Hôpital Nord, Marseilles, France) were screened for inclusion. All trauma patients with an injury severity score greater than 15, who fit the criteria of systemic inflammatory response syndrome, and were between the ages of 18 and 60 yr were invited to participate in the study. They were included after informed consent was signed by their next of kin. Exclusion criteria were pregnancy, immunodeficiency, high-dose corticotherapy (more than 20 mg prednisone or equivalent), or expected death within the first 48 h.

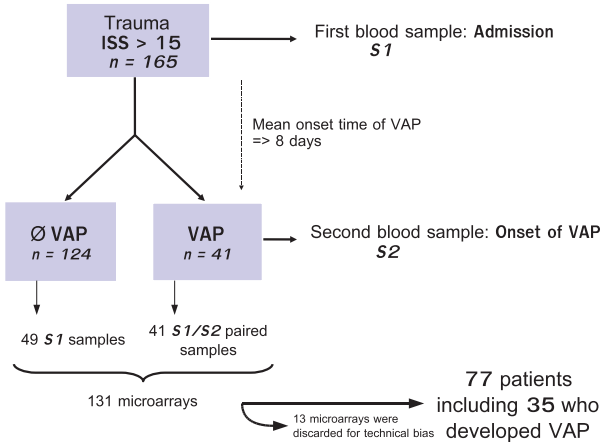


Fig. 1. Flow chart. Samples were collected in 165 trauma patients at admission (S1). In the 41 patients who developed ventilator-associated pneumonia (VAP), a second sample was collected (S2). ISS = injury severity score.

Overall Study Design

The diagnosis of VAP was established when the following criteria were fulfilled: purulent bronchial sputum; body temperature more than 38°C or less than 36°C; leukocytes more than $10 \cdot 10^{-9} \cdot l^{-1}$ or less than $4 \cdot 10^{-9} \cdot l^{-1}$; chest radiograph showing new or progressive infiltrates; presence of at least one microorganism at a concentration of $\geq 10^4$ colony-forming units (cfu)/ml in bronchoalveolar lavage, or $\geq 10^6$ cfu/ml in tracheal aspirates.[20] Senior intensivists and microbiologists made the final diagnosis of VAP. AQ: 1

We defined two sample time points (fig. 1). A blood sample was collected within the first 24 h after ICU admission in all trauma patients fulfilling inclusion criteria (S1). Those who developed VAP were sampled a second time at the onset of antibiotic administration (S2). One hundred thirty-one samples (from 90 patients) were run on microarrays, including 49 S1 samples from patients who did not develop VAP and 41 S1-S2 pairs of samples from patients who developed VAP. Because of technical difficulties, some microarrays were discarded. Therefore, data from 77 patients were analyzed, including 35 paired S1-S2 samples from patients who developed VAP, and 42 S1 samples from patients who did not develop VAP. To exclude a generalized effect related to the elapse of time after trauma, 15 trauma patients who did not develop VAP were sampled between days 5 and 10. Because of a lack of RNA, blood was sampled in 13 additional trauma patients who developed VAP. These 28 samples were used to confirm microarray results. We checked that the two groups of patients analyzed by either microarray (n = 90) or NanoString (n = 28) were similar to the whole population included in the study (see Supplemental Digital Content 1, a table summarizing patients' characteristics for each subgroup, http://links.lww.com/ALN/A756). F1

Sample and Data Collection

Whole blood samples were collected within 24 h of ICU admission (Paxgene®-PreAnalytiX®; Qiagen/Becton Dickson, Valencia, CA) and at the onset of VAP (fig. 1). Blood was withdrawn from an arterial line directly into the PaxGene® tubes. Mixed samples were stored 8 h at 4°C, 24 h at −18°C, and were then maintained at −80°C until RNA extraction. Clinical and biologic data were collected from patient charts at the time of sample collections using our local web-based database.

RNA Extraction and Labeling

Total RNA was isolated from whole blood samples using the PaxGene® Blood RNA System (PreAnalytiX®) according to the manufacturer's specifications. RNA quality was analyzed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), which analyzes the entire RNA electrophoretic trace to evaluate RNA integrity. RNA was extracted from three independent tubes for each time point. High-quality samples were pooled (for each patient and time point). RNA was quantified by spectrophotometry (NanoDrop®; ThermoFisher Scientific, Waltham, MA). As reported previously, hybridization from whole blood samples collected with the PaxGene® system results in a low percentage of genes termed "present" in the microarray analysis. This was mostly attributable to the high expression levels of a small number of genes in the erythrocytes.[21] To improve the hybridization signals, 15 μg each RNA sample, instead of the 5 μg recommended in the standard protocol, was reverse-transcribed into double-stranded complementary DNA (cDNA) and labeled with P^{33}-cytosine.[22] For each time point, all RNAs of each patient that were of good quality, based on the peak profiles of 18S and 28S ribosomal RNA, were pooled. All labeled cDNA generated from these messenger RNA (mRNA) were of the expected radioactive intensity and were hybridized to the microarrays.

Microarray Design and Hybridization

The microarray (HuSG9 k) was developed in our laboratory (Technologie Avancées pour le Génome et la Clinique [TAGC], INSERM U928, Marseilles, France) as described elsewhere.[22] Briefly, HuSG9 k is a human cDNA library, designed from IMAGE clones, which is arranged on a nylon microarray. The microarray contains 9,216 probes comprising 8,682 genes, 100 duplicated genes, and 434 controls. The microarray design was not enriched in specific pathways. Only genes with known functions were selected. Probes were selected by bioinformatics analysis to match the 3′ end of the major transcripts of each gene (Microarray Quality Control, TAGC, INSERM U928). Polymerase chain reaction (PCR) amplification was performed as described previously,[22] and PCR products were spotted onto nylon membranes (Hybond-N+; GE Healthcare, Little Chalfont, Buckinghamshire, England) with a MicroGrid II arrayer (Affymetrix, Santa Clara, CA). PCR amplification products of selected clones were 1,500 base pairs long on average. The robot used a print head with a 16 × 4 arrangement of print tips. Each tip printed 144 spots in a 12 × 12 grid. Additional details regarding the HuSG9 k microarray are available on the TAGC Web site.§§ Because of the high abundance of erythrocyte-related mRNA species in the samples, certain probes were overexpressed and interfered with analysis of the neighboring probes on the microarray. These overly abundant probes were diluted before being spotted on the microarrays. cDNA preparations, hybridizations, and washes were carried out as described previously.[22]

Fn2

NanoString nCounter Assay

In several studies, the NanoString nCounter assay has been shown to be a novel, effective alternative to SYBR green real-time PCR.[23,24] In brief, a multiplexed probe library is made with two sequence-specific probes for each gene of interest. The first probe contains a 35- to 50-base pair sequence, complementary to a particular target mRNA, plus a short common sequence that is coupled to an affinity tag. In our study, the affinity tag is biotin. The second probe contains a second 35- to 50-base sequence, complementary to the target mRNA, which is coupled to an RNA-based, color-coded tag that provides the detection signal. The linear order of these differently colored RNA segments creates a unique code for each gene of interest. Unique pairs of capture and reporter probes are constructed to detect transcripts for each gene of interest. All probes are mixed together with total RNA in a single hybridization reaction that proceeds in solution. After each complex is captured on a surface coated with streptavidin, an applied electric field extends and orients each of them in the solution in the same direction. The complexes are then immobilized in an elongated state and imaged. Each target molecule of interest is identified by the color code generated by the ordered fluorescent segments present on the reporter probe. The level of expression is measured by counting the number of codes for each mRNA. A complete description and performance analysis of the NanoString nCounter assay can be found in the seminal article.[25] In our study, 100 ng each sample total RNA was sent to Seattle's NanoString laboratories. Samples were collected from a cohort of 28 patients who were not analyzed by microarray. The raw results were normalized using internal controls provided by the manufacturer. Differences in the normalized counts were assessed with the Wilcoxon rank sum test and were called significant if $P < 0.05$.

Microarray and Statistical Analysis

Microarray image files were captured with a phosphor imager (BAS5000; FujiFilm, Tokyo, Japan). Data S from .INF or .IMG files were extracted with BZScan2 software (TAGC, INSERM U928).[26] Data files were loaded and analyzed with R and Bioconductor,[27] using the NylonArray library devel-

AQ: 2

§§ http://tagc.univ-mrs.fr/. Last accessed on April 12, 2011.

Table 1. Patient Characteristics at Admission

	Patients Who Did Not Develop VAP (n = 124)	Patients Who Developed VAP (n = 41)
Age (yr)	31 [23–41]	30 [22–46]
BMI (kg/m^2)	24 [23–25]	25 [23–26]
Head trauma		
n (%)	104 (84)	38 (92)
AIS	5 [3–5]	5 [4–5]
Chest trauma		
n (%)	57 (46)	28 (68)*
AIS	3 [0–3]	3 [2–4]
Glasgow Coma Scale	7 [4–11]	5 [3–8]
SAPS II	41 [36–51]	47 [43–53]
Injury Severity Score	29 [24–37]	36 [26–43]*
Heart rate (/min)	116 [100–129]	114 [99–124]
Mechanical ventilation n (%)	96 (77)	38 (92)
Body temperature (°C)	37.9 [36.1–38.7]	38.7 [36.2–39.2]
Creatinine (μM)	81 [70–97]	73 [57–90]
Leukocytes (g/l)	11.9 [8.8–14.7]	13.1 [10.4–16.7]
Hemoglobin (g/dl)	10.5 [8.9–12.5]	10.3 [8.7–11.5]
Platelets (g/l)	178 [135–275]	172 [123–218]
Fibrinogen (mM)	2.4 [1.9–3]	4.5 [2.5–6.6]*
Pao_2/Fio_2 ratio	282 [268–345]	218 [195–287]
Plasma lactate (mM)	1.9 [1.4–3.5]	1.6 [1.2–2]
Death n (%)	24 (19)	11 (27)

Data are presented as median [interquartile range] or as absolute count (percentage).

* $P < 0.05$.

AIS = Acute Injury Score; BMI = body mass index; SAPS = Simplified Acute Physiological Score; VAP = ventilator-associated pneumonia.

oped by the TAGC to support BZScan2 files (library available upon request). Raw data were normalized using quantile normalization. Supervised analysis (supervised methods aimed at finding a set of genes whose expression profiles best correlate with a known phenotype) between admission samples (S1) from patients who developed and those who did not develop VAP, and between the S1 and the S2 groups of samples, was conducted using the significance analysis of microarray algorithm[28] with the siggenes library.[29] All statistical analysis used corrections for multiple comparisons (Benjamini and Hochberg[30]). The power of each analysis was estimated using the Bioconductor library sample size and power analysis.[31] CoPub was used as a literature-based key word enrichment tool for the identified genes.[32] Identified key words are presented along with enrichment significance (*P* value from Fisher exact test).

For clinical and biologic variables, quantitative data are presented as the median and interquartile range, and categorical data are presented as an absolute count and percentages. The differences between groups were evaluated with the Wilcoxon rank sum test or the chi-square test as appropriate. All statistical tests were two-tailed.

Results

Characteristics of Patients

A cohort of 165 trauma patients was prospectively included in the study. Of these, 41 (25%) patients developed VAP. All patients had severe multiple trauma involving the head (86%), chest (52%), abdomen (36%), and limbs (48%). At the time of admission, the injury severity score was higher in the trauma patients who developed VAP than in those who did not develop VAP (median [interquartile range]: 36 [26–43] *vs.* 29 [24–37], $P = 0.04$). At the thoracic level, an acute injury score greater than 2 was found in 28 (68%) patients who developed VAP, compared with 57 (46%) patients who did not develop VAP ($P = 0.02$) (table 1). Death occurred in 11 (27%) patients who developed VAP and in 24 (19%) patients who did not develop VAP ($P = 0.42$). Median onset day of pneumonia was day 8. Clinical and laboratory variables of patients who developed VAP are provided in table 2. The clinical and laboratory characteristics of these patients were similar, except for Pao_2, which was lower at the onset of VAP.

No Transcriptional Signature Predicts VAP in Trauma Patients

To identify prognostic markers of VAP, we compared samples collected at the time of ICU admission from trauma patients who developed VAP and those who did not develop VAP. After we excluded hybridizations because of technical difficulties, 39 admission samples from patients who developed VAP were compared with 42 admission samples from patients who did not develop VAP. The expression profile was similar in the trauma patients who developed VAP and

Table 2. Characteristics of Patients Who Developed Ventilator-associated Pneumonia

	Admission (n = 41)	Onset of VAP (n = 41)
Heart rate (/min)	108 [95–124]	114 [99–128]
Body temperature (°C)	38.4 [36.3–39.2]	38.5 [36.1–39.7]
Creatinine (μM)	77 [69–99]	65 [54–75]*
Glycemia (mmol/l)	6.5 [5.8–7.3]	6.4 [5.8–7.3]
Leukocytes (g/l)	13 [10–17]	13 [10–18]
Plasma lactate (mmol/l)	1.8 [1.2–2.6]	1.5 [1.2–1.8]
Pao_2 (mmHg)	135 [96–175]	91 [79–110]*
Fio_2 (%)	48 [40–56]	50 [40–60]
Sao_2 (%)	99 [98–99]	98 [97–99]

Clinical and biological characteristics of the patients who developed a ventilator-associated pneumonia (VAP), at admission and at the onset of VAP. Data are presented as median [interquartile range].

* $P < 0.01$.

those who did not develop VAP. Therefore, no transcriptional signature predicted the onset of VAP in our trauma patients.

Differential Gene Expression between Admission Time and Onset of VAP in Trauma Patients

To identify diagnostic markers of VAP, we compared admission samples from trauma patients who developed VAP with samples collected in the same patients at the onset of VAP. After the exclusion of six hybridizations (because of technical issues), we analyzed 35 paired samples.

The analysis of normalized gene expression identified 207 probes whose expression levels differed between the two sampling times (false discovery rate less than 10%). The heat map shows this signature by representing, according to a color code, the expression value of each gene (*row*) in each sample (*column*) (see Supplemental Digital Content 2, a figure showing the signature as a heat map, http://links.lww.com/ALN/A757). After hierarchical clustering of expression data, all samples but eight (four patients) clustered together according to the time of sampling. Genes clustered into two groups: 137 (66%) were up-regulated and 70 (34%) were down-regulated at the onset of VAP (see Supplemental Digital Content 2, http://links.lww.com/ALN/A757). Literature-based annotation using CoPub identified several enriched molecular functions (see Supplemental Digital Content 3, a table showing the signature's functional annotation, http://links.lww.com/ALN/A758).[32] Thirty-six genes were up-regulated with a ratio above 2 and seven genes were down-regulated with a ratio below 2 (see Supplemental Digital Content 4, a table showing the 43 most modulated genes, http://links.lww.com/ALN/A759).

To further select potential diagnostic markers, we selected up-regulated genes with a false discovery rate less than 1%. The heat map shows a signature of 10 genes (see Supplemental Digital Content 5, a figure showing the concise signature as a heat map, http://links.lww.com/ALN/A760). The fold changes ranged from 1.3 to 4.2. Supplemental Digital Content 6 lists the identified genes, with their known biologic functions, http://links.lww.com/ALN/A761.

Table 3. Inflammatory Modulation between Admission and Onset of VAP

Gene Symbol	Fold Change Microarray VAP	Fold Change NanoString VAP	Fold Change NanoString Control
IL1B	0.80	0.40	0.92
IL1RN	0.60	0.57	0.84
IL6	1.03	—	—
IL6R	0.93	0.77	0.84
HMGB1	0.77	0.86	0.81
TREM1	0.87	0.70	0.90
ICAM1	0.95	0.97	0.88
TNF	—	0.70	1.04
CALCA	1.21	—	—
CALCB	1.23	—	—
CRP	1.01	—	—

Microarray fold changes are presented for patients who developed VAP. NanoString fold changes are presented for patients who developed VAP and controls. Eleven markers were tested. TNF was not tested in microarray. Four genes were not expressed in NanoString nCounter assay: IL6, CALCA, CALCB, and CRP I.

CALCA = calcitonin-related polypeptide alpha; CALCB = calcitonin-related polypeptide beta; CRP = C-reactive protein; HMGB1 = high mobility group box 1; ICAM1 = intercellular adhesion molecule 1; IL1B = interleukin 1 beta; IL1RN = interleukin 1 receptor antagonist; IL6 = interleukin 6; IL6R = interleukin 6 receptor; TNF = tumor necrosis factor; TREM1 = triggering receptor expressed on myeloid cells 1; VAP = ventilator-associated pneumonia.

Table 3 reports the ratio and baseline expression levels of a sample of genes described previously as markers of inflammation or infection. Although not statistically significant, the expression of interleukin (IL)1B, IL1RN, HMGB1, and TREM1 were down-regulated in the trauma patients at the onset of VAP.

Confirmation of the Selected Markers with the NanoString nCounter Technology

The NanoString nCounter technology was used to validate the expression changes of the 10 genes that are potential diagnostic markers of VAP. We used samples from a new cohort of 13 trauma patients who developed VAP. Results with the NanoString nCounter assay confirmed the modulation of five (ALAS2, AHSP, SPARC, PPBP, and PCSK1) of the 10 selected genes (fig. 2A). For these five genes, we found a statistically significant up-regulation in the time between admission and onset of VAP.

To exclude a generalized effect on gene expression related to trauma and not VAP, we analyzed samples from 15 trauma patients who did not develop VAP. These patients were free of infection and antibiotic treatment. These samples

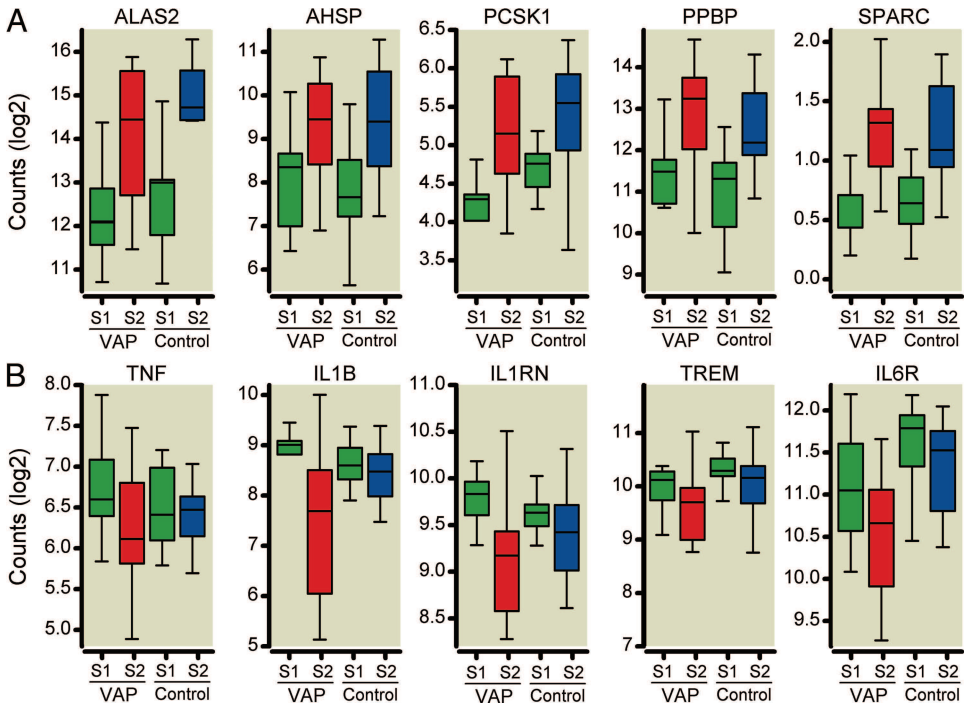


Fig. 2. Confirmation analysis by NanoString nCounter assay in trauma patients who developed ventilator-associated pneumonia (VAP) and trauma patients who did not develop VAP (controls). Patients were sampled at admission (S1). A second sample (S2) was collected in patients who developed VAP (at the onset of VAP). A second sample was collected in controls between days 5 and 10 (S2). Counts in log2 scale for the five confirmed genes of the concise signature (*A*) and five proinflammatory markers (*B*) are presented as box plots (*green* = S1; *red* = S2 in patients with VAP; *blue* = S2 in controls).

were collected between days 5 and 10. The signature of 10 genes was tested in these samples. A similar profile of expression was found, including the five selected genes (fig. 2A).

Decrease of Proinflammatory Gene Expression Is Associated with VAP

We also tested a set of already known inflammatory genes: TREM1, TNF, IL6R, IL6, IL1B, IL1RN, ICAM1, HMGB1, CRP, CALCA, and CALCB. Not enough signal was detected for IL6, CRP, CALCA, and CALCB. These four genes were excluded from the analysis. With the use of NanoString nCounter technology, IL1B, IL1RN, TNF, IL6R, and TREM1 were down-regulated in samples from patients at the onset of VAP. In addition, these five proinflammatory markers differentiated patients who developed VAP and those who did not develop VAP (fig. 2B).

Discussion

Using a whole blood transcriptome analysis, the Sepsi*Chip* project was designed to identify a signature of gene expression associated with the development of VAP. Among 165 multiple trauma patients, 41 developed VAP. As previously described, we confirmed an association between the development of VAP and the severity of trauma.[33] Despite these clinical differences, the transcriptome analysis does not show any difference between trauma patients who did or did not developed VAP. Therefore, this analysis cannot serve to predict the clinical outcome of trauma patients. Comparison of samples at the time of admission and at the time of VAP onset identifies a 207-gene transcriptional signature. These genes are involved in basal cell functions, such as transcription, translation, apoptosis, immune response, hemostasis, and coagulation. Of note, genes involved in the proinflammatory response are nonsignificantly down-regulated. Focusing on genes with a false discovery rate less than 1%, we identified a 10-gene signature for the potential diagnosis of VAP. Using NanoString, we confirmed the up-regulation of 5 of these 10 genes. However, we found a similar signature in a cohort of trauma patients who did not develop VAP, suggesting a generalized effect related to trauma. In an additional analysis, we identified a down-regulation of proinflammatory mediators only in the patients who developed VAP.

In contrast to experimental studies, clinical investigations often are limited by excessive heterogeneity of the included patients.[34] To reduce this bias, we selected trauma patients in a predetermined age range. We also used a single model of infection. Because genetic backgrounds can differ among patients, we constructed a paired analysis. In our study, analysis of samples from the same patient at admission and at onset of VAP serves as a control for genetic heterogeneity. This design also let us compare the host response to inflammation (trauma) and infection (VAP). This may improve the characterization of genes modulated by infection, independently of inflammation. In agreement with the findings of previous

studies,[4,35,36] classic markers of the inflammatory response in our study were up-regulated at admission. This early and transient proinflammatory response was also found in a murine model of acute lung injury.[37] A signature of 207 probes discriminates the samples of patients at the time of admission and at the onset of VAP. Actually, we found that this signature is related to a generalized response after trauma. Thus, no specific signature of infection is found in trauma patients who developed VAP.

This striking finding shows that the transcriptional changes observed in the trauma patients are caused by inflammation and other general effects after trauma. In other studies reporting signatures for severe sepsis, the controls are healthy volunteers.[10,12,38] Thus, these studies cannot differentiate infection and inflammation. The signatures that were reported in these previous studies may be related to inflammation and not specifically to the infectious process.

In a seminal paper, McDunn *et al.* suggest that the onset of an infection-specific transcriptional program may precede the clinical diagnosis of VAP. Using a bench-to-bedside approach, 85 genes were modulated during the 7 days bracketing the diagnosis of VAP. A general trajectory (entitled "riboleukogram") was identified.[6] In our study, we tested 52 of these 85 genes. Forty-one (79%) genes exhibited a similar modulation. In light of our findings, the trajectories associated with the riboleukogram may be related to inflammation more than infection. However, the concept of a general trajectory may be more useful diagnostically than a single-point analysis.

Obtaining high-quality genomic data from critically ill patients remains a challenge. Cell collection methods, mRNA extraction, and labeling methods influence the expression profiles. Buffy coat *versus* whole blood sampling is discussed elsewhere.[8,21] As described previously, whole blood samples resulted in a weak hybridization signal. Because a few probes exhibited overexposed signals in preliminary experiments (data not shown), we changed our protocols by diluting these probes on all microarrays. Next, we increased the quantity of RNA and P^{33}-cytosine for the cDNA preparations. These changes corrected the percentage of present calls into a normal range. These issues are related to the use of whole blood samples. However, both its reproducibility and its simplicity in the setting of a multicenter clinical study drive the choice of Paxgene®.

Our microarray findings were confirmed by using the novel NanoString nCounter Analysis System. NanoString uses digital technology based on direct multiplexed measurement of gene expression. It offers a high level of sensitivity, precision, and reproducibility.[25] NanoStrings are fluorescent bar codes that bind to target mRNA for gene expression analysis. The advantages are its sensitivity comparable with PCR, the lack of enzymology or amplification, and the ability to analyze the expression levels of as many as 800 genes in a single assay. Using this system, we confirm the modulation observed by microarray analysis for 5 of the 10 gene of the signature.

To exclude a general effect after trauma, we tested these five potential biomarkers in a set of patients who did not develop VAP. Indeed, the inflammatory profile of trauma patients is affected by different procedures, such as mechanical ventilation, parenteral nutrition, and surgical procedures. In this cohort, the five genes are up-regulated, despite that these patients did not develop VAP. Thus, these genes are not specific for an infectious process. We also tested a set of markers that are classically involved in inflammation and infection. Five proinflammatory mediators are down-regulated in the patients who developed VAP, whereas they are unchanged in our controls. The same down-regulation is also detected in the microarray data.

This down-regulation in patients who develop VAP is of interest because the polarization of the immune response in ICU patients remains a matter of debate. This supports the concept of immunosuppression in ICU patients with infection.[39] As highlighted elsewhere,[40] the immune system of septic patients is unable to respond to a second challenge of endotoxin. This state of nonresponsiveness is known as "anergy."[41] Autopsy studies in persons who died of sepsis showed a profound, apoptosis-induced loss of adaptive immune system cells.[40] Notably, most studies aimed at inhibiting proinflammatory cytokines failed to improve survival. In contrast, stimulating the immune response in selected immunodeficient septic patients, by granulocyte/macrophage-colony stimulating factor, for example, may improve their outcome.[42] In the current study, we found a decreased expression of TNF, IL1B, IL1RN, IL6, and TREM1 in the patients who developed VAP. Therefore, we confirm the association between a relative immunodeficiency and the sepsis occurrence. This suggests the need to monitor the immunologic status of trauma patients.

Our study has several limitations. It is possible that the heterogeneity of patients included in the study precludes the identification of a positive effect. However, our patients were composed of a relatively homogeneous population of patients with multiple traumas, including 80% with a head injury. To reduce the effect of age, our inclusion criteria were restrictive. Our endpoint was restricted to the development of VAP. Except for seven positive urine cultures, for which the rate of occurrence was similar in patients who did or did not develop VAP, no infection interfered with VAP. The initial analysis, designed to compare the transcriptome at the time of admission for trauma patients who did or did not develop VAP, was negative. Even with a high false discovery rate (20%), we did not find any differentially expressed genes. Thus, we interpreted this result as the absence of prognostic biomarkers of VAP in trauma patients. We also conducted subgroup analysis to infer the potential effect of sex, age, or other infection on transcriptome modulation. No effect was found, but these results (data not shown) should be interpreted with caution because of low power estimates.

Because the blood can be described as a black box,[43] one can suggest that it may not reflect the transcriptional modulation in the lung. Analysis of whole blood from trauma patients most likely does not reflect transcriptional changes in the lung. To address this, we have studied time-related changes in the lung using an animal model of lung injury.[37]

Conclusion

In conclusion, we explored the transcriptome of whole blood samples from 165 trauma patients, 41 of whom developed VAP. Our goal was to identify a gene expression profile that could be used to identify trauma patients at risk for VAP. At the time of admission, the trauma patients who developed VAP had a higher severity score than did those who did not develop VAP. In addition, we found that chest trauma is a risk factor for VAP. Within the limitations of our study, for patients that developed VAP, a transcriptional signature was identified that reflected differential gene expression at time of admission and time of VAP diagnosis. This signature is attributed to a generalized pattern related to trauma and seems independent of the infectious process. Interestingly, five proinflammatory markers are down-regulated at the onset of VAP, although the differences are not statistically significant. They are expressed at a steady level in the patients who did not develop VAP. These results support the concept that infection in the ICU results from an unbalanced shift toward an antiinflammatory state. The trajectory of the inflammatory balance in trauma patients at clinical risk of VAP should be assessed in future studies.

This work is dedicated to Antoine Secondi (chief nursing officer), our friend. Antoine taught us to care, a job in which he excelled; he died the day before this work was accepted. The authors are indebted to Dominique Lemoine, Frédéric Navas, Véronique Paone, Marie-Hélène Po, Antoine Secondi (Certified Registered Nurse Anesthetists, Hôpital Nord, Assistance Publique - Hôpitaux de Marseille, Marseille) for the management of samples; Khadija Hammi (Clinical Research Assistant, Assistance Publique - Hôpitaux de Marseille, Marseille) and Nathalie Lesavre, M.D. (Centre d'Investigation Clinique, Hôpital Nord, Marseille) for case report form completion and administrative tasks; and François Antonini, M.D. (Service d'Anesthésie et de Réanimation, Hôpital Nord, Assistance Publique - Hôpitaux de Marseille, Université de la Méditerranée, Marseille) and Pascal Rihet, Ph.D. (U958, Université de la Méditerranée, Marseille) for useful comments on the statistical approach.

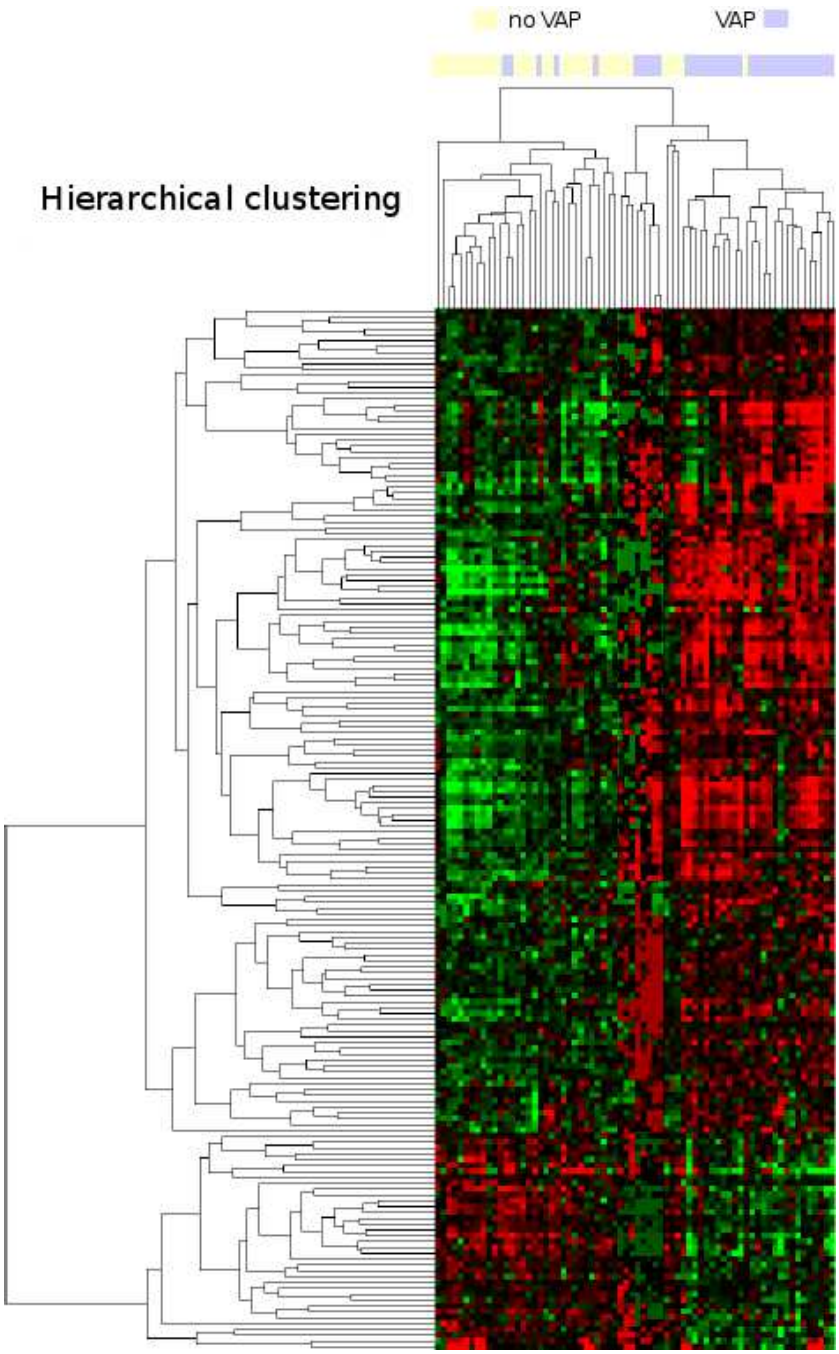
References

1. Antonelli M, Moro ML, Capelli O, De Blasi RA, D'Errico RR, Conti G, Bufi M, Gasparetto A: Risk factors for early onset pneumonia in trauma patients. Chest 1994; 105:224–8
2. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, Suppes R, Feinstein D, Zanotti S, Taiberg L, Gurka D, Kumar A, Cheang M: Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med 2006; 34: 1589–96
3. Croce MA, Swanson JM, Magnotti LJ, Claridge JA, Weinberg JA, Wood GC, Boucher BA, Fabian TC: The futility of the clinical pulmonary infection score in trauma patients. J Trauma 2006; 60:523–7
4. Maier M, Wutzler S, Lehnert M, Szermutzky M, Wyen H, Bingold T, Henrich D, Walcher F, Marzi I: Serum procalcitonin levels in patients with multiple injuries including visceral trauma. J Trauma 2009; 66:243–9
5. Pelosi P, Barassi A, Severgnini P, Gomiero B, Finazzi S, Merlini G, d'Eril GM, Chiaranda M, Niederman MS: Prognostic role of clinical and laboratory criteria to identify early ventilator-associated pneumonia in brain injury. Chest 2008; 134:101–8
6. McDunn JE, Husain KD, Polpitiya AD, Burykin A, Ruan J, Li Q, Schierding W, Lin N, Dixon D, Zhang W, Coopersmith CM, Dunne WM, Colonna M, Ghosh BK, Cobb JP: Plasticity of the systemic inflammatory response to acute infection during critical illness: Development of the riboleukogram. PLoS One 2008; 3:e1564
7. Maier M, Wutzler S, Bauer M, Trendafilov P, Henrich D, Marzi I: Altered gene expression patterns in dendritic cells after severe trauma: Implications for systemic inflammation and organ injury. Shock 2008; 30:344–51
8. Cobb JP, Mindrinos MN, Miller-Graziano C, Calvano SE, Baker HV, Xiao W, Laudanski K, Brownstein BH, Elson CM, Hayden DL, Herndon DN, Lowry SF, Maier RV, Schoenfeld DA, Moldawer LL, Davis RW, Tompkins RG, Baker HV, Bankey P, Billiar T, Brownstein BH, Calvano SE, Camp D, Chaudry I, Cobb JP, Davis RW, Elson CM, Freeman B, Gamelli R, Gibran N, Harbrecht B, Hayden DL, Heagy W, Heimbach D, Herndon DN, Horton J, Hunt J, Laudanski K, Lederer J, Lowry SF, Maier RV, Mannick J, McKinley B, Miller-Graziano C, Mindrinos MN, Minei J, Moldawer LL, Moore E, Moore F, Munford R, Nathens A, O'keefe G, Purdue G, Rahme L, Remick D, Sailors M, Schoenfeld DA, Shapiro M, Silver G, Smith R, Stephanopoulos G, Stormo G, Tompkins RG, Toner M, Warren S, West M, Wolfe S, Xiao W, Young V: Application of genome-wide expression analysis to human health and disease. Proc Natl Acad Sci USA 2005; 102:4801–6
9. Laudanski K, Miller-Graziano C, Xiao W, Mindrinos MN, Richards DR, De A, Moldawer LL, Maier RV, Bankey P, Baker HV, Brownstein BH, Cobb JP, Calvano SE, Davis RW, Tompkins RG: Cell-specific expression and pathway analyses reveal alterations in trauma-related human T cell and monocyte pathways. Proc Natl Acad Sci USA 2006; 103:15564–9
10. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, Miller-Graziano C, Moldawer LL, Mindrinos MN, Davis RW, Tompkins RG, Lowry SF: A network-based analysis of systemic inflammation in humans. Nature 2005; 437:1032–7
11. Schaecher K, Kumar S, Yadava A, Vahey M, Ockenhouse CF: Genome-wide expression profiling in malaria infection reveals transcriptional changes associated with lethal and nonlethal outcomes. Infect Immun 2005; 73:6091–100
12. Shanley TP, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Doctor A, Kalyanaraman M, Tofil NM, Penfil S, Monaco M, Odoms K, Barnes M, Sakthivel B, Aronow BJ, Wong HR: Genome-level longitudinal expression of signaling pathways and gene networks in pediatric septic shock. Mol Med 2007; 13:495–508
13. Ramilo O, Allman W, Chung W, Mejias A, Ardura M, Glaser C, Wittkowski KM, Piqueras B, Banchereau J, Palucka AK, Chaussabel D: Gene expression patterns in blood leukocytes discriminate patients with acute infections. Blood 2007; 109: 2066–77
14. Prucha M, Ruryk A, Boriss H, Möller E, Zazula R, Herold I, Claus RA, Reinhart KA, Deigner P, Russwurm S: Expression profiling: Toward an application in sepsis diagnostics. Shock 2004; 22:29–33
15. Payen D, Lukaszewicz AC, Belikova I, Faivre V, Gelin C, Russwurm S, Launay JM, Sevenet N: Gene profiling in human blood leucocytes during recovery from septic shock. Intensive Care Med 2008; 34:1371–6
16. Pachot A, Lepape A, Vey S, Bienvenu J, Mougin B, Monneret G: Systemic transcriptional analysis in survivor and non-

survivor septic shock patients: A preliminary study. Immunol Lett 2006; 106:63–71

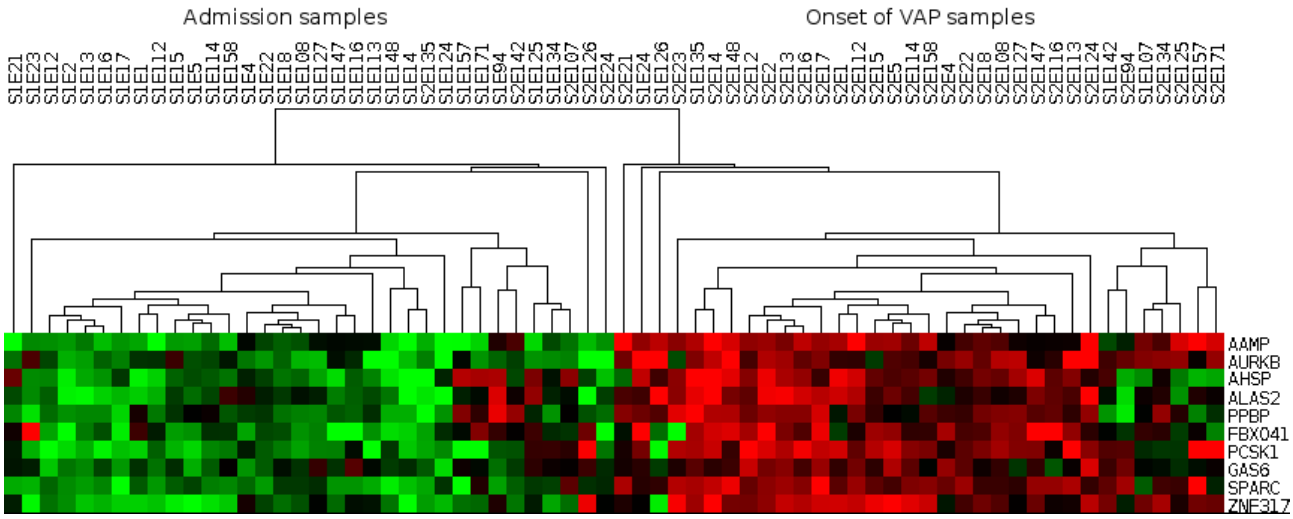
17. Johnson SB, Lissauer M, Bochicchio GV, Moore R, Cross AS, Scalea TM: Gene expression profiles differentiate between sterile SIRS and early sepsis. Ann Surg 2007; 245:611–21
18. Feezor RJ, Oberholzer C, Baker HV, Novick D, Rubinstein M, Moldawer LL, Pribble J, Souza S, Dinarello CA, Ertel W, Oberholzer A: Molecular characterization of the acute inflammatory response to infections with gram-negative *versus* gram-positive bacteria. Infect Immun 2003; 71:5803–13
19. Tang BM, McLean AS, Dawes IW, Huang SJ, Cowley MJ, Lin RC: Gene-expression profiling of gram-positive and gram-negative sepsis in critically ill patients. Crit Care Med 2008; 36:1125–8
20. American Thoracic Society, Infectious Diseases Society of America: Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. Am J Respir Crit Care Med 2005; 171:388–416
21. Feezor RJ, Baker HV, Mindrinos M, Hayden D, Tannahill CL, Brownstein BH, Fay A, MacMillan S, Laramie J, Xiao W, Moldawer LL, Cobb JP, Laudanski K, Miller-Graziano CL, Maier RV, Schoenfeld D, Davis RW, Tompkins RG, Inflammation and Host Response to Injury, Large-Scale Collaborative Research Program: Whole blood and leukocyte RNA isolation for gene expression analyses. Physiol Genomics 2004; 19:247–54
22. Puthier D, Joly F, Irla M, Saade M, Victorero G, Loriod B, Nguyen C: A general survey of thymocyte differentiation by transcriptional analysis of knockout mouse models. J Immunol 2004; 173:6109–18
23. Payton JE, Grieselhuber NR, Chang LW, Murakami M, Geiss GK, Link DC, Nagarajan R, Watson MA, Ley TJ: High throughput digital quantification of mRNA abundance in primary human acute myeloid leukemia samples. J Clin Invest 2009; 119:1714–26
24. Malkov VA, Serikawa KA, Balantac N, Watters J, Geiss G, Mashadi-Hossein A, Fare T: Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System. BMC Res Notes 2009; 2:80
25. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, James JJ, Maysuria M, Mitton JD, Oliveri P, Osborn JL, Peng T, Ratcliffe AL, Webster PJ, Davidson EH, Hood L, Dimitrov K: Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 2008; 26:317–25
26. Lopez F, Rougemont J, Loriod B, Bourgeois A, Loï L, Bertucci F, Hingamp P, Houlgatte R, Granjeaud S: Feature extraction and signal processing for nylon DNA microarrays. BMC Genomics 2004; 5:38
27. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J: Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol 2004; 5:R80
28. Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001; 98:5116–21
29. Schwender H, Ickstadt K: Empirical Bayes analysis of single nucleotide polymorphisms. BMC Bioinformatics 2008; 9:144
30. Benjamini Y, Hochberg Y: Controlling the false discovery rate, a practical and powerful approach to multiple testing. J R Statist Soc B 1995; 57:289–300 AQ: 3
31. Ferreira JA, Zwinderman A: Approximate power and sample size calculations with the Benjamini-Hochberg method. Int J Biostat 2006; 2:8
32. Frijters R, Heupers B, van Beek P, Bouwhuis M, van Schaik R, de Vlieg J, Polman J, Alkema W: CoPub: A literature-based keyword enrichment tool for microarray data analysis. Nucleic Acids Res 2008; 36:W406–10
33. Leone M, Brégeon F, Antonini F, Chaumoître K, Charvet A, Ban LH, Jammes Y, Albanèse J, Martin C: Long-term outcome in chest trauma. ANESTHESIOLOGY 2008; 109:864–71
34. Singer M: Mitochondrial function in sepsis: Acute phase *versus* multiple organ failure. Crit Care Med 2007; 35:S441–8
35. Sauerland S, Hensler T, Bouillon B, Rixen D, Raum MR, Andermahr J, Neugebauer EA: Plasma levels of procalcitonin and neopterin in multiple trauma patients with or without brain injury. J Neurotrauma 2003; 20:953–60
36. Martin C, Boisson C, Haccoun M, Thomachot L, Mege JL: Patterns of cytokine evolution (tumor necrosis factor-alpha and interleukin-6) after septic shock, hemorrhagic shock, and severe trauma. Crit Care Med 1997; 25:1813–9
37. Lesur I, Textoris J, Loriod B, Courbon C, Garcia S, Leone M, Nguyen C: Gene expression profiles characterize inflammation stages in the acute lung injury in mice. PLoS One 2010; 5:e11485
38. Wong HR, Shanley TP, Sakthivel B, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Doctor A, Kalyanaraman M, Tofil NM, Penfil S, Monaco M, Tagavilla MA, Odoms K, Dunsmore K, Barnes M, Aronow BJ, Genomics of Pediatric SIRS/Septic Shock Investigators: Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome. Physiol Genomics 2007; 30:146–55
39. Hotchkiss RS, Coopersmith CM, McDunn JE, Ferguson TA: The sepsis seesaw: Tilting toward immunosuppression. Nat Med 2009; 15:496–7
40. Hotchkiss RS, Karl IE: The pathophysiology and treatment of sepsis. N Engl J Med 2003; 348:138–50
41. Munoz C, Carlet J, Fitting C, Misset B, Blériot JP, Cavaillon JM: Dysregulation of *in vitro* cytokine production by monocytes during sepsis. J Clin Invest 1991; 88:1747–54
42. Meisel C, Schefold JC, Pschowski R, Baumann T, Hetzger K, Gregor J, Weber-Carstens S, Hasper D, Keh D, Zuckermann H, Reinke P, Volk HD: Granulocyte-macrophage colony-stimulating factor to reverse sepsis-associated immunosuppression: A double-blind, randomized, placebo-controlled multicenter trial. Am J Respir Crit Care Med 2009; 180:640–8
43. Leone M, Boutière B, Camoin-Jau L, Albanèse J, Horschowsky N, Mège JL, Martin C, Dignat-George F: Systemic endothelial activation is greater in septic than in traumatic-hemorrhagic shock but does not correlate with endothelial activation in skin biopsies. Crit Care Med 2002; 30:808–14

Supplemental Figure 1: Transcriptional signature in trauma patients who developed ventilator-associated pneumonia (VAP).



Heatmap representation of the 207 identified genes with a false discovery rate below 10%. Samples were classified according to sampling time (admission: yellow; onset of VAP: blue). Four patients were misclassified. Genes are represented as rows, and samples as columns. The level of gene expression is color-coded from green (under-expression) to red (over-expression).

Supplemental Figure 2: Concise transcriptional signature in trauma patients who developed ventilator-associated pneumonia (VAP).



Heatmap representation of the 10 identified genes with a false discovery rate below 1%. Samples were classified according to sampling time. Five patients were misclassified. Genes are represented as rows, and samples as columns. The level of gene expression is color-coded from green (under-expression) to red (over-expression).

Table S1: Patient characteristics according to type of analysis

	All population *n = 165*	**Patients analyzed by microarray** *n = 90*	**Patients analyzed by NanoString** *n = 28*
Age (year)	31 [23-42]	30 [22-45]	29 [21-43]
SAPS II	43 [38-52]	45 [40-52]	41 [36-51]
Injury Severity Score	30 [24-39]	33 [26-41]	30 [24-40]
Head trauma n (%)	142 (86)	79 (88)	23 (82)
Chest trauma n (%)	85 (52)	54 (60)	15 (54)
Death n (%)	35 (21)	21 (23)	5 (18)

Comparison of general characteristics for the whole population, patients analyzed by microarray, and patients analyzed by NanoString nCounter. SAPS : Simplified Acute Physiological Score. Data are presented as median [inter-quartile range] or absolute count (percentage).

Table S2: Signature genes' functionnal annotation

Keyword	***p-value***	**Number of genes**
Growth	1.3x10-5	84
Transcription	6.9x10-5	88
Translation	7.0x10-5	50
Hemostasis	2.6x10-4	12
Innate immune response	5.2x10-4	19
Apoptosis	9.2x10-4	60
Immune response	4.8x10-3	31
Coagulation	5.8x10-3	12

Main keywords, extracted from the literature, associated with the signature's genes with number of genes in the signature, and corresponding p-value (Fisher exact test).

Table S3: Most differentially expressed genes

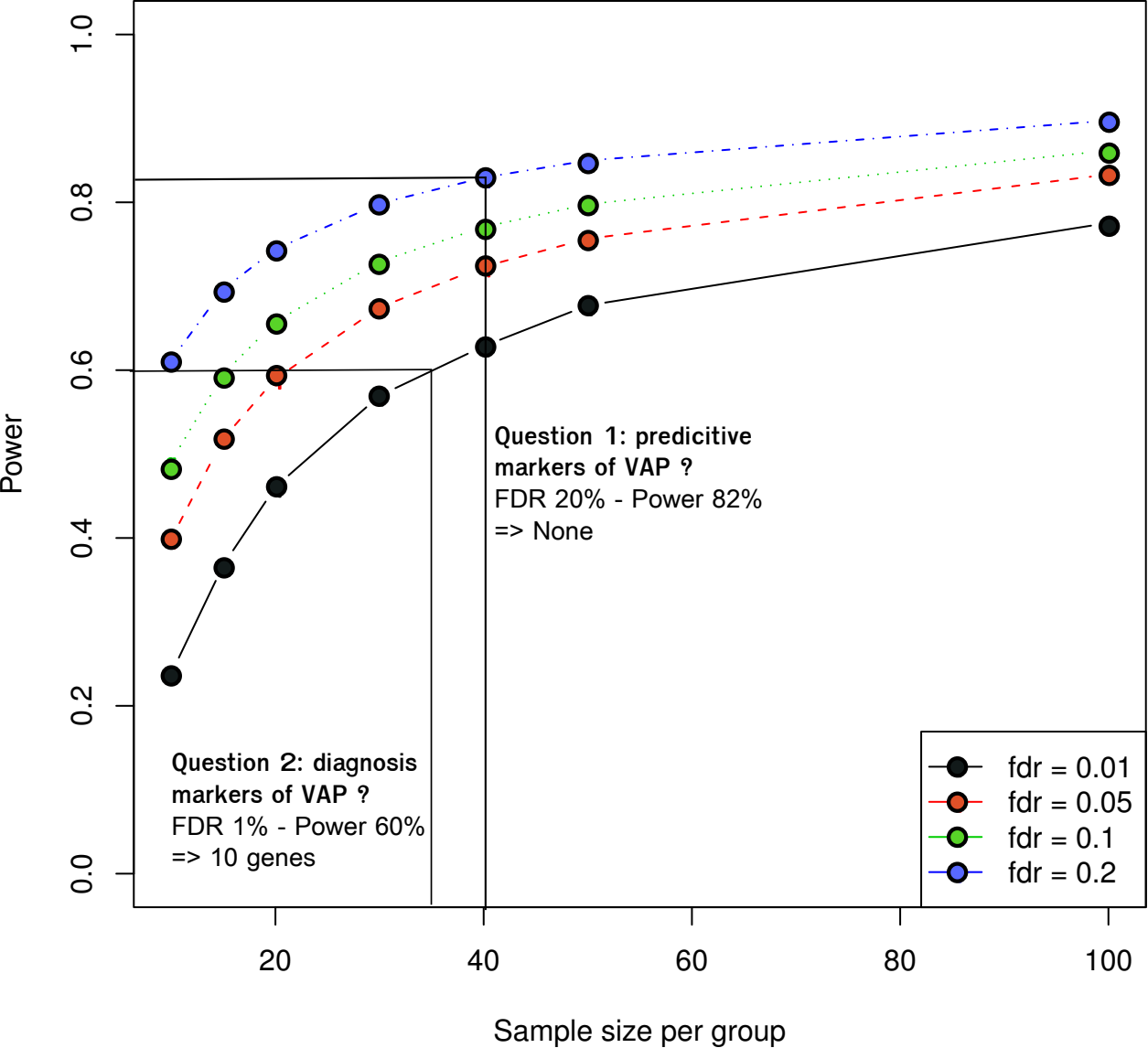
Up-regulated genes		Down-regulated genes	
Gene symbol	**Fold change**	**Gene symbol**	**Fold change**
PCSK1	4.2	LPIN1	-2.4
AAMP	4.1	ORF1-FL49	-2.4
SPARC	3.5	ACSL1	-2.4
ZNF317	3.4	CD3G	-2.2
IGHG3	3.1	ZDHHC19	-2.2
DEFA4	3.0	CHDH	-2.1
AURKB	2.9	PVM1	-2.1
AHSP	2.8		
PPBP	2.7		
NARG1	2.7		
BPGM	2.7		
PRSS25	2.6		
TREML1	2.5		
KHDRBS2	2.4		
EIF2AK1	2.4		
SELENBP1	2.4		
LCN2	2.3		
GTF2E2	2.3		
FBXO41	2.3		
NRGN	2.3		
IGHA1	2.2		
TMCC2	2.2		
IGKV1-5	2.2		
SFRS3	2.1		
ITGA10	2.1		
TENS1	2.1		
FOXO3A	2.1		
RAB40A	2.0		
ALAS2	2.0		
P2RY5	2.0		
PF4	2.0		
CSDA	2.0		
HDGF	2.0		
RPL35A	2.0		
STK19	2.0		
EPB42	2.0		

Genes differentially expressed between admission and onset of ventilator-associated pneumonia (VAP) in trauma patients who developed VAP. Gene symbols are provided according to fold changes. Only genes with an absolute fold change above 2 are presented.

Table S4: Concise signature genes' functions.

Symbol	Function (from UNIPROT)	Fold change
GAS6	Ligand for tyrosine-protein kinase receptor AXL, TYRO3 and MER whose signaling is implicated in cell growth and survival, cell adhesion and cell migration. Plays a role in thrombosis by amplifying platelet aggregation and secretion in response to known agonists.	**1.3**
ALAS2	Erythroid specific 5-aminolevulinate synthase. Catalyzes Succinyl-CoA + Glycine <=> 5-aminolevulinate + CoA + CO2.	**2.0**
FBXO41	Substrate-recognition component of the SCF (SKP1-CUL1-F-box protein)-type E3 ubiquitin ligase complex	**2.3**
PPBP	LA-PF4 stimulates DNA synthesis, mitosis, glycolysis, intracellular cAMP accumulation, prostaglandin E2 secretion, and synthesis of hyaluronic acid and sulfated glycosaminoglycan. It also stimulates the formation and secretion of plasminogen activator by human synovial cells. NAP-2 is a ligand for CXCR1 and CXCR2, and NAP-2, NAP-2(73), NAP-2(74), NAP-2(1-66), and most potent NAP-2(1-63) are chemoattractants and activators for neutrophils. TC-1 and TC-2 are antibacterial proteins, in vitro released from activated platelet alpha-granules. CTAP-III(1-81) is more potent than CTAP-III desensitize chemokine-induced neutrophil activation.	**2.7**
AHSP	Acts as a chaperone to prevent the harmful aggregation of alpha-hemoglobin during normal erythroid cell development. Specifically protects free alpha-hemoglobin from precipitation. It is predicted to modulate pathological states of alpha-hemoglobin excess such as beta-thalassemia	**2.8**
AURKB	May be directly involved in regulating the cleavage of polar spindle microtubules and is a key regulator for the onset of cytokinesis during mitosis. Component of the chromosomal passenger complex (CPC), a complex that acts as a key regulator of mitosis. The CPC complex has essential functions at the centromere in ensuring correct chromosome alignment and segregation and is required for chromatin-induced microtubule stabilization and spindle assembly. Phosphorylates 'Ser-10' and 'Ser-28' of histone H3 during mitosis. Required for kinetochore localization of BUB1 and SGOL1.	**2.9**
ZNF317	May function as a transcription factor. May play an important role in erythroid maturation and lymhoid proliferation.	**3.4**
SPARC	Appears to regulate cell growth through interactions with the extracellular matrix and cytokines. Binds calcium and copper, several types of collagen, albumin, thrombospondin, PDGF and cell membranes. There are two calcium binding sites; an acidic domain that binds 5 to 8 Ca2+,with a low affinity and an EF-hand loop that binds a Ca2+ ion with a high affinity.	**3.5**
AAMP	Plays a role in angiogenesis and cell migration. In smooth muscle cell migration, may act through the RhoA pathway.	**4.1**
PCSK1	Involved in the processing of hormone and other protein precursors at sites comprised of pairs of basic amino acid residues. Substrates include POMC, renin, enkephalin, dynorphin somatostatin and insulin.	**4.2**

Genes function. Known function extracted from UNIPROT for each of the ten identified genes with the concise signature.


Power
1.0
0.8
0.6
0.4
0.2
0.0
20
40
60
80
100
Sample size per group
Question 1: predicitive markers of VAP ?
FDR 20% - Power 82%
=> None
Question 2: diagnosis markers of VAP ?
FDR 1% - Power 60%
=> 10 genes
fdr = 0.01
fdr = 0.05
fdr = 0.1
fdr = 0.2

Sex-Related Differences in Gene Expression Following Coxiella burnetii Infection in Mice: Potential Role of Circadian Rhythm

Julien Textoris, Leang Heng Ban, Christian Capo, Didier Raoult, Marc Leone, Jean-Louis Mege
PLoS ONE 5(8): e12190

La fièvre Q est la forme clinique de l'infection par *Coxiella burnetii*. Cette zoonose présente un dimorphisme sexuel ; les hommes sont atteints de manière plus fréquente et plus sévère que les femmes, bien que l'exposition soit identique. Cette étude explore le rôle du sexe dans un modèle murin d'infection par *C. burnetii*.

Des souris ont été infectées par *C. burnetii* pendant 24 h, puis le transcriptome hépatique a été analysé par puces à ADN. Une analyse multiclasse a permis d'identifier 2 777 sondes dont l'expression est modulée par l'infection à *C. burnetii.* Seuls 14 % des gènes présentaient une modulation de leur expression indépendante du sexe, alors que 86 % des gènes étaient modulés spécifiquement chez les mâles et les femelles. Une analyse similaire conduite chez des animaux castrés montre que les hormones sexuelles influencent plus de 60 % de la modulation induite par l'infection. La réduction de modulation de l'expression était prépondérante chez les mâles castrés. L'annotation fonctionnelle des gènes modulés a identifié quatre clusters fonctionnels chez les mâles, reliés à l'adhésion inter-cellulaire, la transduction du signal, les défensines et la voie Jak/Stat. L'induction précoce de la transcription de l'IL10 et de Stat3 pourrait expliquer la plus forte susceptibilité des hommes infectés par *C. burnetii* à développer des formes cliniques. Deux clusters fonctionnels ont été identifiés chez les femelles, dont le contrôle du rythme circadien, constitué de plusieurs facteurs de transcription interagissant au niveau d'une boucle de rétro-contrôle. L'expression de Clock et d'Arntl (bras inducteur de la boucle) est diminuée alors que l'expression de Per (bras inhibiteur de la boucle) est augmentée. Cette modulation précoce pourrait être associée à une élimination plus efficace des bactéries chez les souris femelles par rapport aux mâles chez lesquels l'infection est plus marquée.

Cette étude confirme au niveau transcriptionnel l'existence d'un dimorphisme sexuel dans l'infection par *C. burnetii*. Elle montre pour la première fois que le rythme circadien joue un rôle important dans la réponse anti-infectieuse chez la souris, et pose de nouvelles bases pour élucider le rôle du dimorphisme sexuel en pathologie infectieuse humaine.

OPEN ACCESS Freely available online

PLoS one

Sex-Related Differences in Gene Expression Following *Coxiella burnetii* Infection in Mice: Potential Role of Circadian Rhythm

Julien Textoris[1,2], Leang Heng Ban[1,2], Christian Capo[1], Didier Raoult[1], Marc Leone[1,2], Jean-Louis Mege[1]*

1 Unité de Recherche sur les Maladies Infectieuses Tropicales et Emergentes, Centre National de la Recherche Scientifique Unité Mixte de Recherche 6236, Faculté de Médecine, Marseille, France, **2** Service d'anesthésie et de réanimation, Hôpital Nord, Assistance Publique – Hôpitaux de Marseille, Université de la Méditerranée, Marseille, France

Abstract

Background: Q fever, a zoonosis due to *Coxiella burnetii* infection, exhibits sexual dimorphism; men are affected more frequently and severely than women for a given exposure. Here we explore whether the severity of *C. burnetii* infection in mice is related to differences in male and female gene expression profiles.

Methodology/Principal Findings: Mice were infected with *C. burnetii* for 24 hours, and gene expression was measured in liver cells using microarrays. Multiclass analysis identified 2,777 probes for which expression was specifically modulated by *C. burnetti* infection. Only 14% of the modulated genes were sex-independent, and the remaining 86% were differentially expressed in males and females. Castration of males and females showed that sex hormones were responsible for more than 60% of the observed gene modulation, and this reduction was most pronounced in males. Using functional annotation of modulated genes, we identified four clusters enriched in males that were related to cell-cell adhesion, signal transduction, defensins and cytokine/Jak-Stat pathways. Up-regulation of the IL-10 and Stat-3 genes may account for the high susceptibility of men with Q fever to *C. burnetii* infection and autoantibody production. Two clusters were identified in females, including the circadian rhythm pathway, which consists of positive (Clock, Arntl) and negative (Per) limbs of a feedback loop. We found that Clock and Arntl were down-modulated whereas Per was up-regulated; these changes may be associated with efficient bacterial elimination in females but not in males, in which an exacerbated host response would be prominent.

Conclusion: This large-scale study revealed for the first time that circadian rhythm plays a major role in the anti-infectious response of mice, and it provides a new basis for elucidating the role of sexual dimorphism in human infections.

Citation: Textoris J, Ban LH, Capo C, Raoult D, Leone M, et al. (2010) Sex-Related Differences in Gene Expression Following *Coxiella burnetii* Infection in Mice: Potential Role of Circadian Rhythm. PLoS ONE 5(8): e12190. doi:10.1371/journal.pone.0012190

Editor: Frank R. DeLeo, National Institute of Allergy and Infectious Diseases, National Institutes of Health, United States of America

Received April 27, 2010; **Accepted** July 22, 2010; **Published** August 13, 2010

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Jean-Louis.Mege@univmed.fr

Introduction

Social factors such as gender inequity can explain differences in the distribution of infectious diseases between men and women. As shown elsewhere, poor women may be at a disadvantage in their ability to access quality health care [1]. However, biological differences are also responsible for part of the epidemiological variation observed between males and females in infectious diseases due to intra- and extracellular pathogens [2]. Gender-based biological differences also affect host immune responses to pathogens. Women elicit more vigorous humoral and cell-mediated immune responses than men in response to antigenic challenges, while men have frequently been observed to exhibit more aggressive and harmful inflammatory responses to pathogens [3]. Tuberculosis [4,5] and Legionnaire's disease [6] are reported to be more prevalent and severe in men than in women. Although biological differences have been largely attributed to sex hormones [7], the precise nature of the cross-talk between sex and infection remains largely unknown.

Q fever is a zoonosis due to *Coxiella burnetii*, an intracellular bacterium [8,9]. Following primary infection, half of infected patients experience acute Q fever. The disease is characterized by clinical polymorphism and includes fever, granulomatous hepatitis, and pneumonia. For a similar exposure risk, men are 2.5 times more likely to become symptomatic than women [10]. Hyperinflammatory Q fever with granulomatous hepatitis and autoantibodies has been reported in male patients who respond poorly to antimicrobial agents [11]. In addition, Q fever complications are higher in males than in females. As a result, males represent 75% of patients diagnosed with *C. burnetii* endocarditis [12]. In mice, the severity of *C. burnetii* infection is also sex-dependent, with females exhibiting lower bacterial load than males. Ovariectomy increases the bacterial load in the liver and spleen, and this is corrected by 17β-estradiol treatment [13], demonstrating that estrogens limit tissue infection. Nevertheless, the underlying mechanisms that govern the differential responsiveness of males and females to bacterial infection are poorly understood. To investigate these differences, we adopted a large-scale approach

consisting of a microarray covering the whole genome. Unexpectedly, 86% (2,379/2,777) of the probes that were modulated by *C. burnetii* infection were dependent on sex. We identified a specific pathway related to the circadian cycle in females that may control the severity of *C. burnetii* infections.

Results and Discussion

Mice were infected with *C. burnetii* for 24 hours, and changes in gene expression were investigated in liver cells. *C. burnetii* infection generated more transcriptional changes in males than in females. When samples were plotted according to the weight of variances due to sex and infection, principal component analysis (PCA) clearly discriminated males and females and revealed that the distance between uninfected and infected males was higher than that found between uninfected and infected females (Fig. 1A). A multiclass analysis identified 2,777 probes that were modulated in response to *C. burnetii*. These probes were classified into four distinct clusters (Fig. 1B). The first cluster is sex-independent and includes 398 probes (14% of the total) that were similarly regulated in both males and females. The other three clusters, representing the majority of probes modulated by *C. burnetii* infection (86% of the total), are sex-dependent: 1,459 probes (53%) were specifically modulated in males whereas 892 (32%) were specifically modulated in females. The fourth cluster of 28 probes is divergent between males and females, with 14 probes up-regulated in males and 14 up-regulated in females (for the list, see Table S1). Using *in silico* analysis, we related the modulated genes to the subcellular distribution of their encoded proteins (Fig. 2). In infected males, the modulated genes encode proteins that have a high degree of interconnection. These interactions spread across the membrane, underlying that a profound reorganization occurred at the membrane after infection. In infected females, there were less interactions between proteins encoded by modulated genes at the membrane level than in males, but the interconnection between the cytosol and the nucleus was apparently more developed. These results highlight important changes in transcription in male and female mice infected with *C. burnetii*.

It is known that nonsexual tissues and cells manifest sex-related differences in most animals including humans. Many of these differences are dependent on sex hormones but it is seems difficult to attribute other differences to hormones [14]. A microarray approach performed on long-lived mice has showed that 381 genes (on over 14,000 genes) are altered in these mice compared to wild type mice in a sex-independent manner while 110 genes are affected in a sex-dependent manner [15]. Our analysis take into account genes modulated in response to *C. burnetii*: these genes are included in two categories, namely sex-independent and sex-dependent genes. Unexpectedly, sex-dependent genes represent 86% of the total number of genes modulated in response to *C. burnetii*. Among them, the majority (80%) were not modulated upon sex in uninfected animals. It is well known that, in humans, the clinical expression of different infectious diseases including tuberculosis [5], Legionnaire's disease [6] and Q fever [10] is associated to sex. Sex has been often described as involved in the incidence of common infectious diseases such as pneumonia [16]. Similar results have been obtained in mice infected with *Listeria monocytogenes* [17] or *Plasmodium berghei* [18]. Bacterial sepsis, a major cause of morbidity and mortality in humans, is due to an exaggerated response of hosts [19]. The frequency and the severity of bacterial sepsis are also associated with sexual dimorphism [2,20]. Interestingly, such differences are also found in isolated human circulating leukocytes [21] and animal models of endotoxemia [22]. Finally, only few data show that sex does not influence human infections: this has been reported for cerebrospinal [23] and soft tissue [24] infections.

The molecular mechanisms that underlie the differences between males and females in bacterial infections are only in part

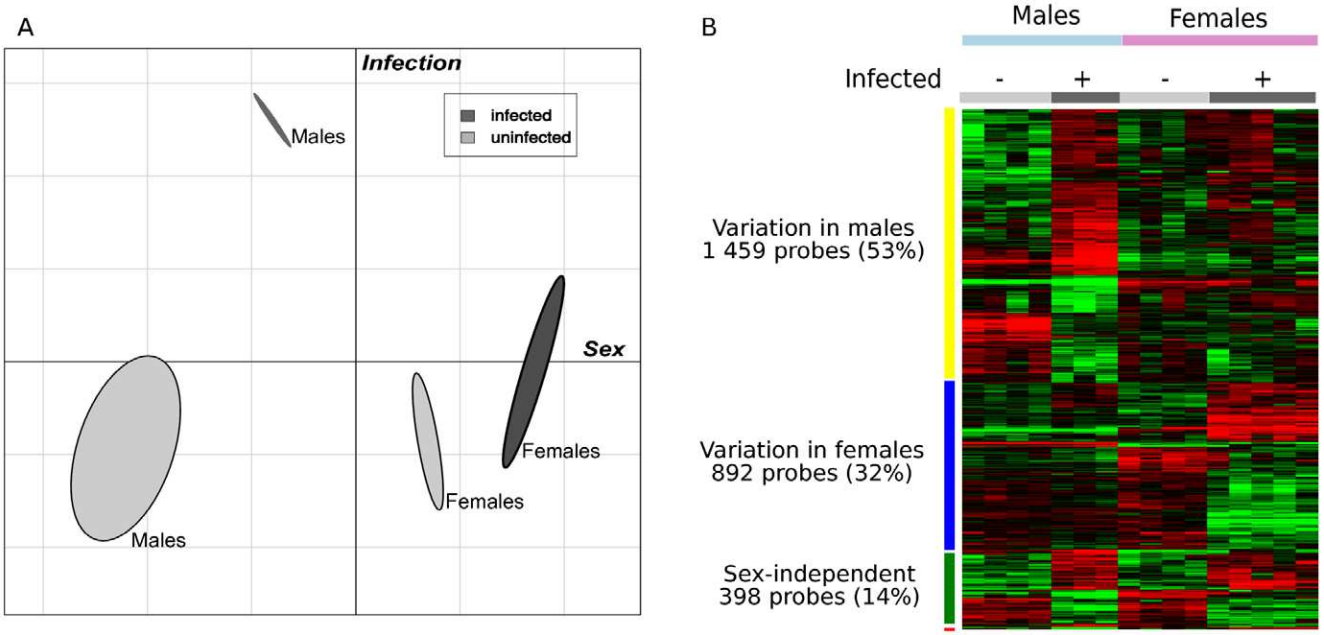


Figure 1. Impact of sex on *C. burnetii* infection. Male and female mice were infected with *C. burnetii* for 24 hours. Transcriptional responses were assessed by microarray. A, The impacts of sex and infection on gene expression were analyzed by principal component analysis using R. Each axis distance represents the amount of variance in gene expression explained by the corresponding factor (sex or infection). B, Hierarchical clustering analysis was used to classify selected up-regulated and down-regulated genes in four clusters: sex-independent genes, male-dependent genes, female-dependent genes and genes inversely modulated in males and females (red vertical bar).
doi:10.1371/journal.pone.0012190.g001

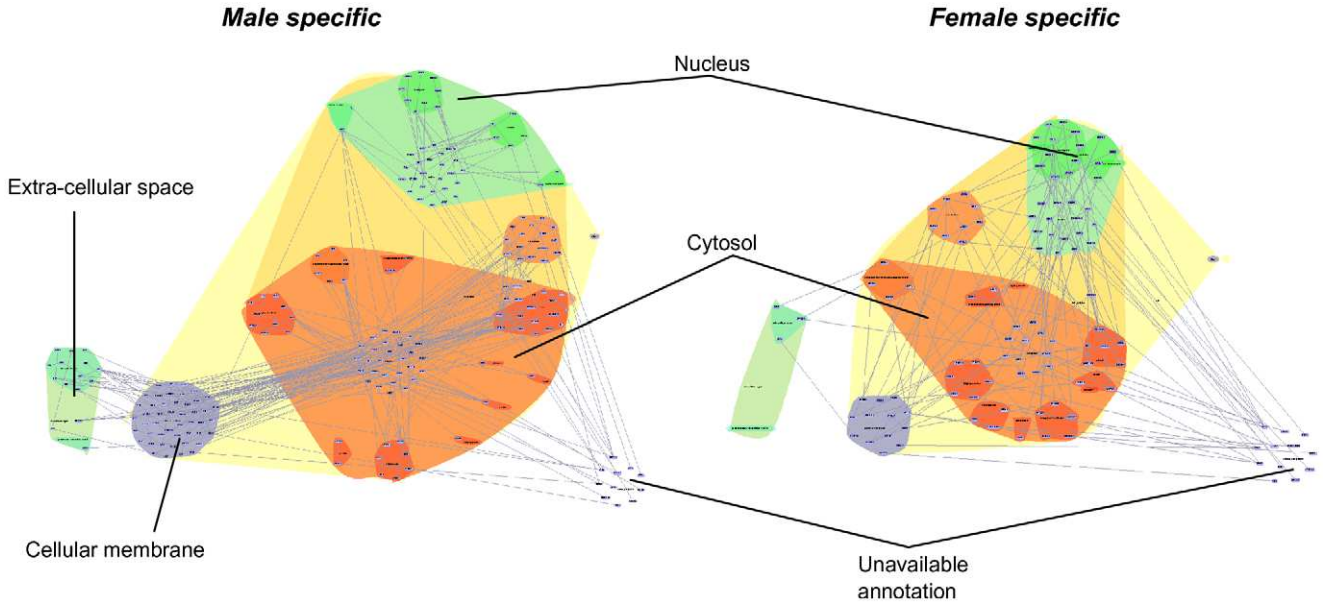


Figure 2. Subcellular distribution of proteins encoded by modulated genes. The InteractomeBrowser plugin of TranscriptomeBrowser was used to generate a graph of protein interactions in males and female mice according to their cellular distribution [65].
doi:10.1371/journal.pone.0012190.g002

characterized. As sexual hormones may play a major role in the transcriptional responses to *C. burnetii* infection in male and female mice, we investigated the effect of castration on these responses. Castration shortened the distance between uninfected and infected males, reflecting less pronounced changes in gene expression (Fig. S1A). Castration did not substantially modify the impact of infection on females (Fig. S1B). After *C. burnetii* infection, the number of modulated probes designated as sex-independent remained remarkably constant in castrated mice (396 vs. 398 in intact mice). In contrast, among the sex-dependent probes, only 902 probes were modulated in castrated mice compared with 2,379 probes in intact mice, suggesting that sex hormones were responsible for more than 60% of the gene modulation (Fig. S2). This reduction was most significant in males: 378 probes were specifically modulated in castrated males (vs. 1,459 in intact males) compared to 516 in castrated females (vs. 892 in intact females), suggesting that testosterone was responsible for the major changes observed after *C. burnetii* infection. This finding is in line with previous studies suggesting the predominant effect of testosterone on sexual dimorphism [25,26]. It is important to note that about 40% of sex-dependent genes that were modulated in response to *C. burnetii* were not linked to hormonal status. This is reminiscent of previous data that showed that sex differences are not completely abrogated or accentuated by hormone ablation or supplementation, respectively [27].

We next used functional annotation of the modulated genes to identify major pathways that were possibly involved in the response to *C. burnetii* infection. In the sex-independent cluster, four groups of enriched keywords were identified (Table 1). Group 1, with an enrichment score of 4.3, is related to early induction of the acute phase liver response. Acute phase response is a major function of the liver that contributes to bacterial clearance in acute Q fever [10]. The gene encoding serum amyloid A1–3 was expressed with a fold change (FC) of over 25. The genes encoding orosomucoid 1–3, serine peptidase inhibitor and complement components (C8a, C9) were also up-regulated. Group 2 (enrichment score of 3.9) is related to lipid metabolism and group 3

Table 1. Functional annotation of the sex-independent cluster.

Ontology	Keywords	*P*-value
Acute phase response (ES : 4.3–25 genes) SAA1, SAA2, SAA3, SAA4, ORM3, SERPINA3N, C8a, C9		
SP_PIR_KEYWORDS	Acute phase	2.9×10^{-6}
GO_BP	Acute-phase response	6.4×10^{-6}
GO_BP	Acute inflammatory response	1.8×10^{-5}
INTERPRO	Serum amyloid A protein	8.6×10^{-5}
SMART	SAA	6.1×10^{-6}
Lipid metabolism (ES : 3.9–38 genes) CRAT, ACOX1, ACOT2, ACOT5, EHHADH, ACAA1A, CPT1B, PLTP		
GO_BP	Fatty acid metabolic process	1.3×10^{-7}
GO_BP	Monocarboxylic acid metabolic process	9.1×10^{-9}
SP_PIR_KEYWORDS	Fatty acid metabolism	8.4×10^{-6}
BO_BP	Lipid metabolic process	1.6×10^{-9}
GO_CC	Peroxysome	3.3×10^{-5}
GO_CC	Microbody	3.3×10^{-5}
Steroid metabolic process (ES : 3.3–14 genes) STAR, IDI1, MOGAT1, LEPR, FABP5, PCTP, AKR1C18, FBP2, HMGCS1		
GO_BP	Steroid metabolic process	1.7×10^{-4}
GO_BP	Alcohol metabolic process	5.5×10^{-4}
GO_BP	Cholesterol metabolic process	6.8×10^{-4}
GO_BP	Sterol metabolic process	1.0×10^{-3}
Thrombospondin (ES : 1.8–5 genes) SPON2, C9, SEMA5B, ADAMTS20, C8A		
SMART	TSP1	1.7×10^{-3}
INTERPRO	Thrombospondin, type I	1.9×10^{-3}

The enrichment score (ES), the number of genes and example genes in each group of genes were indicated.
doi:10.1371/journal.pone.0012190.t001

(enrichment score of 3.3) is related to steroid metabolism. Most of the genes annotated with keywords of groups 2 and 3 were down-regulated by a mean FC of 2. This finding is apparently related to disturbed lipid metabolism during *C. burnetii* infection since pronounced lipid infiltration and increased cholesterol have been reported in the livers of *C. burnetii*-infected guinea pigs [28]. In addition, *C. burnetii* replication requires free access to host cholesterol stores [29]. Group 4 is related to genes encoding proteins with a thrombospondin domain. Thrombospondin is a family of multifunctional proteins involved in coagulation, angiogenesis, apoptosis, and immune regulation [30]. One of these genes (Spon2) is a pathogen-associated molecular pattern recognition molecule known to recognize *Escherichia coli* lipopolysaccharide and numerous gram negative bacteria [31]. Although some partners of thrombospondin such as $\alpha v\beta 3$ integrin have been described as *C. burnetii* receptors [32], there is no evidence that Spon2 recognizes *C. burnetii*.

Functional annotation of genes that were modulated only in infected males identified four groups of keywords (Table 2). Groups 1 and 3 are related to signal transduction and cell-cell adhesion, respectively. Keyword group 2 is related to genes encoding defensins. Although defensins are peptides with antimicrobial properties [33,34], their role against *C. burnetii* infection has never been reported. Defensins such as Defcr3, Hamp1 and Hamp2 were down-regulated in infected males, suggesting that the reduced expression of defensins favors early infection in males. Hepcidins (Hamp1 and Hamp2) reduce iron availability for invading microorganisms by reducing extracellular iron concentrations [35]. As has been recently suggested, greater iron availability in males may favor *C. burnetii* replication [36]. It has also been shown that hepcidins are differentially expressed in male and female mice [37,38]. Keyword group 4 is related to cytokines and the Jak/Stat pathway. The gene encoding interleukin (IL)-10 was up-regulated in males. These results were confirmed by RT-PCR (Table S2), and a time course study showed that males overproduced interleukin (IL)-10 mRNA compared with females independently of the duration of infection (Fig. 3A). Note that the expression of the gene encoding Stat 3, a transcription factor known to be targeted by IL-10, was higher in males than in females. We previously showed that IL-10 enables monocytes to support *C. burnetii* replication [39,40] and that chronic evolution of Q fever is associated with increased production of IL-10 [41]. However, to our knowledge, a link between IL-10 and sexual dimorphism in *C. burnetii* infection has never been reported. Interferon (IFN)-γ, which is critical for protection against intracellular bacteria and *C. burnetii* infection [42], was weakly modulated in male and female mice one day after infection as determined by microarray and RT-PCR (Table S2). A time course analysis revealed that the gene encoding IFN-γ was dramatically up-regulated at days 4 and 7 post-infection (Fig. 3B), suggesting that IFN-γ induces deleterious effects in parallel with the resolution of *C. burnetii* infection. The expression of the IL-6 gene was slightly up-regulated in females (Table S2), and a time course study showed that the IL-6 gene was dramatically over-expressed at days 4 and 7 post-infection, especially in females (Fig. 3C). After trauma-hemorrhage, IL-6 release by macrophages is suppressed in male mice but not in female mice [43], suggesting that sex and IL-6 production are related. Taken together, these results suggest that males develop an early anti-inflammatory response that may favor increased bacterial burden at the onset of infection and may promote antibody production in an IL-10-dependent manner. This may account for the presence of auto-antibodies in men with Q fever.

The specific signature of infected females, which includes 892 genes (Fig. 1B), was characterized by two major groups of keywords (Table 3): intracellular location and transcriptional activity, and unexpectedly, circadian rhythm. The latter pathway was further investigated with RT-PCR and network analysis. Circadian rhythm is controlled by an auto-regulated transcription-translation feedback loop that regulates the expression of rhythmic genes in a tissue-specific manner [44]. In liver cells, a significant number of clock-regulated genes are associated with the cell cycle and proliferation. Clock and Arntl form the positive limb of the transcriptional loop, while the Per and Cry protein families form the negative limb of the feedback loop by inhibiting their own Clock/Arntl-induced transcription [44]. The turnover of Per and Cry allows the cycle to begin again (Fig. S3). Microarray and RT-PCR data showed that Clock and Arntl expression were down-regulated in females infected with *C. burnetii* for one day while Per expression was up-regulated (Table S2). We questioned whether these variations were persistent because we had previously found that *C. burnetii* infection of liver cells peaks after four days and decreases thereafter [13]. The levels of Arntl and Clock transcripts remained constant in males over seven days, but in females, these

Table 2. Functional annotation of male-dependent cluster.

Ontology	Keywords	*P*-value
Signal transduction (ES : 9.9–418 genes) EXTL1, CDH11, RHO, ADRA2B, IL12RB2, CD36 GLP2R, OLFR981		
SP_PIR_KEYWORDS	Transmembrane	1.2×10^{-9}
GO_BP	Signal transduction	3.9×10^{-7}
GO_MF	Transmembrane receptor activity	1.7×10^{-9}
GO_BP	G-protein coupled receptor signal transduction	4.1×10^{-7}
INTERPRO	Olfactory receptor	8.8×10^{-8}
GO_GC	Transmembrane region	2.0×10^{-6}
Defensin (ES : 8.2–45 genes) DEFB3, DEFA1, DEFCR7, DEFCR17, DEFCR9, DEFCR25, HAMP2		
SMART	DEFSN	5.7×10^{-11}
INTERPRO	Mammalian defensin	1.7×10^{-10}
GO_BP	Defense response to bacterium	2.2×10^{-8}
SP_PIR_KEYWORDS	Antimicrobial	9.4×10^{-7}
GO_BP	Defense response	1.4×10^{-3}
SP_PIR_KEYWORDS	Defensin	1.2×10^{-6}
Cell-cell adhesion (ES : 3.7–52 genes) CDH1, CDH4, FAIM2, MMD2, PCDHA5, PCDHB19, CDH11		
INTERPRO	Cadherin	2.9×10^{-7}
SP_PIR_KEYWORDS	Cell adhesion	2.8×10^{-4}
GO_BP	Cell adhesion	2.5×10^{-3}
SP_PIR_KEYWORDS	Calcium	3.0×10^{-3}
Cytokines/Jak-STAT pathway (ES : 2.3–28 genes) IFNA1, IL5, PDGFA, IL1RN, STAT3, EGF,CXCL12, IL10, IL11, IFNG, IL13		
BIOCARTA	Cytokines and Inflammatory response	2.6×10^{-4}
KEGG_PATHWAY	Jak-STAT signalling pathway	1.7×10^{-3}
KEGG_PATHWAY	Cytokin-Cytokin receptor interaction	4.2×10^{-3}
INTERPRO	Four-helical cytokine, core	7.1×10^{-3}
SP_PIR_KEYWORDS	Cytokine	2.1×10^{-2}
GO_MF	Cytokine activity	2.0×10^{-2}

The enrichment score (ES), the number of genes and example genes in each group of genes were indicated.
doi:10.1371/journal.pone.0012190.t002

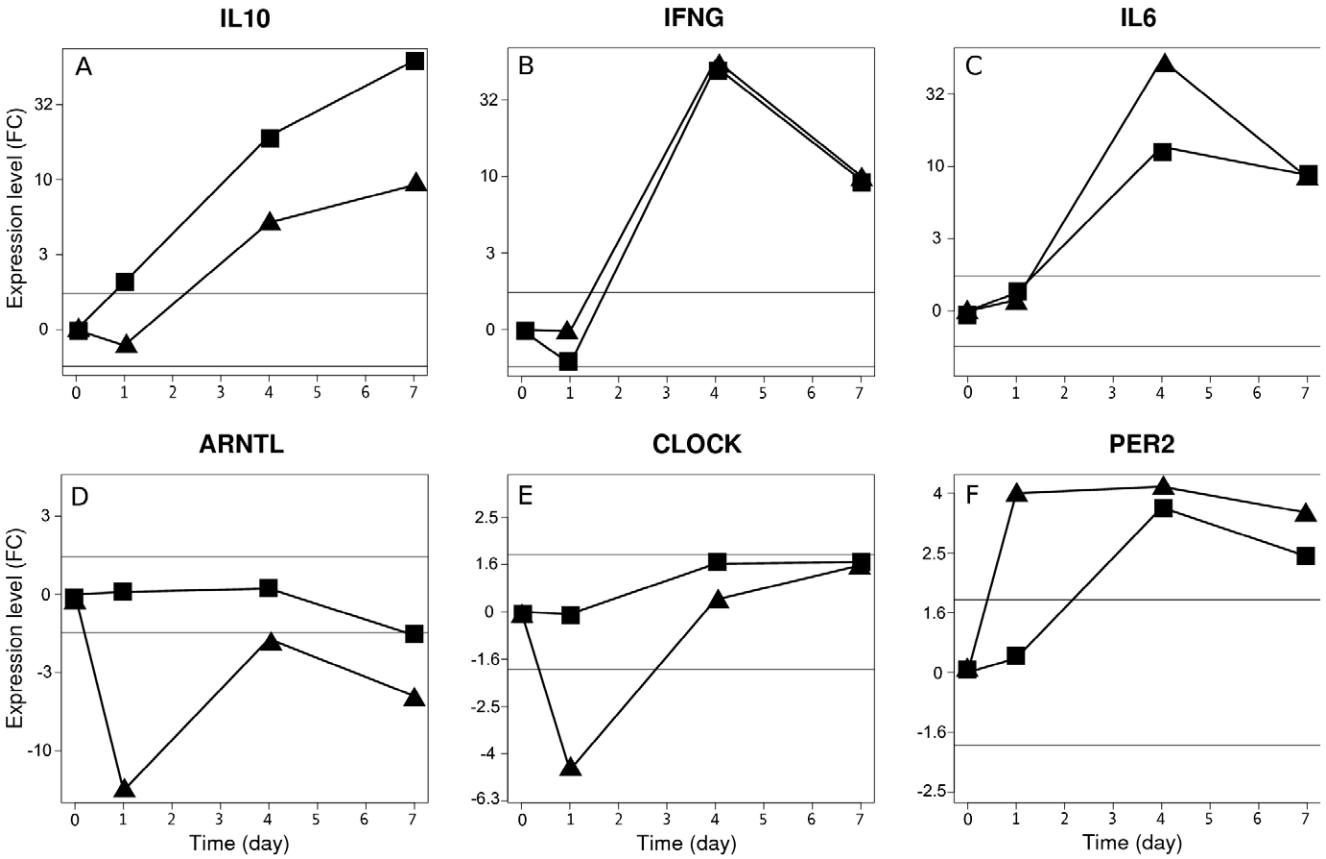


Figure 3. Time course of gene modulation. Male (squares) and female (triangles) mice were infected with *C. burnetii* for one, four and seven days, and gene expression was determined by qRT-PCR and normalized with the GAPDH gene. A–C, genes encoding cytokines. D–F, genes involved in circadian rhythm. Horizontal lines indicate the selected fold change cut-off of ±1.75.
doi:10.1371/journal.pone.0012190.g003

transcripts declined at day one and increased thereafter to approach the number observed in males (Fig. 3D and 3E, respectively). The expression of Per2, which increased in females at day one, remained high during the seven days of the infection. In males, the expression of the Per2 gene increased at day four (Fig. 3F). The differences between males and females regarding the expression of Arntl, Clock and Per2 were lost after castration of the mice (Fig. S4). Taken together, these findings show that *C. burnetii* infection affected the circadian rhythm of female mice. These results are distinct from those shown in a recent report in which the expression of the Clock, Arntl and Per genes was down-modulated in blood leukocytes after the administration of endotoxin to human volunteers. However, the relationship between these changes and sexual dimorphism was not investigated [45].

Table 3. Functional annotation of female-dependent cluster.

Ontology	Keywords	*P*-value
Circadian rhythm (ES : 2.8–14 genes) CLOCK, PER1, PER2, PER3, PTGDS, OPN3, MTNR1A, ARNTL, AFP		
GO_BP	Circadian rhythm	2.9×10^{-8}
GO_BP	Rhythmic process	1.0×10^{-6}
SP_PIR_KEYWORDS	Biological rhythms	2.0×10^{-6}
KEGG_PATHWAY	Circadian rhythm	3.6×10^{-5}
INTERPRO	PAS domain	8.5×10^{-4}
Intracellular/Transcription (ES : 2.2–328 genes) SOX9, PBEF1, STX1B1, RNF20, UGT1A2		
GO_CC	Intracellular	3.5×10^{-5}
GO_CC	Membrane-bound organelle	2.5×10^{-3}
GO_CC	Nucleus	7.3×10^{-3}
INTERPRO	Basic helix-loop-helix dimerization region	2.0×10^{-3}
GO_BP	Metabolic process	2.8×10^{-4}

The enrichment score (ES), the number of genes and example genes in each group of genes were indicated.
doi:10.1371/journal.pone.0012190.t003

The circadian rhythm modification associated with *C. burnetti* infection may affect numerous biological functions. The biological clock controls the secretion of estrogen hormones [46] by regulating the intracellular levels of estrogen receptors α and β [47,48]. Lower Per2 levels lead to the stabilization of estrogen receptor β. As Per2 is estrogen-inducible, a feedback mechanism may attenuate stimulation by estrogens [48]. Male steroid hormones retroactively play a role in the circadian rhythm at the brain level [49]. Orchidectomy leads to a loss of androgen receptor expression and circadian-specific locomotor activity, both of which are restored by testosterone replacement [49,50]. Interestingly, circadian rhythm and immune responses are interconnected [51]. For instance, the susceptibility or resistance of flies to infection varies according to the time when they are infected [52]. The circadian oscillation of clock genes affects IFN-γ

production through the regulation of Per2 [53] or the activity of natural killer cells [54]. Circadian rhythm may also control the steroid pathway since it post-translationally inhibits glucocorticoid receptors [55]. Cross-talk between these pathways may reinforce the effect of sex on *C. burnetii* infection. Finally, we hypothesize that an early response involving circadian rhythm likely promotes efficient bacterial elimination in females, while in males, the early phase of the response to *C. burnetii* infection is inefficient, due at least in part to an anti-inflammatory response that favors bacterial growth [13].

In conclusion, different transcriptional responses to *C. burnetii* were associated with the severity of the infection in male and female mice. Functional annotation showed that the modulated genes were organized in different networks in males and females and we hypothesize that circadian rhythm may be involved in *C. burnetii* infection. As the percentage of sex-dependent genes that were modulated by *C. burnetii* infection was dramatically high, one can expect that numerous genes would be modulated in other infectious diseases described as epidemiologically associated with sexual dimorphism. Our data provide a new basis for elucidating the role of sexual dimorphism in human infections. It would be also interesting to determine if differences in gene expression occur in infectious diseases not described as epidemiologically sex-dependent.

Materials and Methods

Infection of Mice

The following experimental protocol was approved by the Institutional Animal Care and Use Committee of the Université de la Méditerranée. Forty C57BL/6 mice (20 males and 20 females) were obtained from Charles River Laboratories. Ten females and ten males were sterilized at seven weeks of age. Ovaries were removed through bilateral incisions of the lumbar region and testes were removed through bilateral scrotum incision. After three weeks, ten males and ten females, including five non-castrated and five castrated animals of each gender, were killed. The other series of mice were injected intraperitoneally with 10^5 *C. burnetii* organisms at the same time and sacrificed at day one after infection [56]. In another set of experiments, 16 mice (4 intact males, 4 castrated males, 4 intact females and 4 castrated females) were killed at zero, one, four and seven days after *C. burnetii* infection. Organs were aseptically excised. Samples for RNA extraction were stabilized in RNA*later* (Qiagen, France). All specimens were stored at -80°C until use.

RNA isolation and real-time PCR

Total RNA extraction and reverse transcription were performed as previously described [57]. The primers were designed by Primer3 (v. 0.4.0) available at http://frodo.wi.mit.edu/. Reverse trancriptase was omitted in negative controls. The FC in target gene cDNA relative to endogenous control (glyceraldehyde 3-phosphate dehydrogenase, GAPDH) was determined with the formula: fold change $= 2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct = (Ct_{Target} - Ct_{GAPDH})_{infected\ mice} - (Ct_{Target} - Ct_{GAPDH})_{uninfected\ mice}$. Ct values were defined as the cycle number at which fluorescence signals were detected [58].

Microarray procedures and data analysis

Whole Mouse Genome Oligo Microarray 4×44K Kit (44,000 60-mer oligonucleotides) and One-Color Microarray-Based Gene Expression Analysis (Agilent Technologies) were used as recently described [57]. Slides were washed, dried, and scanned at 5-µm resolution with a G2505B DNA microarray scanner (Agilent Technologies). Intensity signals were generated using the Agilent Feature Extractor Software A.9.1.3 after image quality control was performed manually. Four aberrant samples (two infected males, one uninfected male and one uninfected female) were eliminated before starting the analysis because of technical problems. Raw signal data were normalized with the quantile method [59] and then transformed into binary logarithm. The Significance Analysis of Microarrays test (SAM [60]) was used to study the gene expression in intact animals with a multiclass analysis (sex and infection). False Discovery Rate (FDR) was set to 0.05 to filter modulated genes. Then, selected probes (n = 13,814) were filtered using the FC obtained from dividing the median intensity of infected and non-infected mice. FCs of male and female mice were calculated with a cut-off at 1.75 as selection criterium for differentially expressed genes. Thus, the selected genes which were annotated as "sex-specific" were also "infection-specific". Supervised analyses were carried out with R (R for statistical computation and graphic version 2.7.2, GPL) with the library BioConductor [61]. Functional annotations and classifications were performed using the DAVID Bioinformatics Resource 2008 [62]. Keywords from the following ontologies were analyzed (GO, SP_PIR_KEYWORDS, UP_SEQ_FEATURE, SMART, INTERPRO, KEGG_PATHWAYS, and BIOCARTA). Gene Ontology (GO) keywords are further classified according to their main category: Biological Process (GO_BP), Cellular Component (GO_CC) or Molecular Function (GO_MF). PCA was used to visually explore global effects for genome wide trends, unexpected effects and outliers in the expression data (library made4 for R [63]). The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [64] and are accessible through GEO Series accession number GSE21065 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21065).

Supporting Information

Figure S1 Impact of castration on *C. burnetii* infection. Male and female mice were castrated and then infected with *C. burnetii* for 24 hours. Transcriptional responses were assessed by microarray. The impacts of castration and infection on gene expression were assessed by principal component analysis in males (A) and females (B) using R. Each axis distance represents the amount of variance in gene expression explained by the corresponding factor (castration or infection).
Found at: doi:10.1371/journal.pone.0012190.s001 (7.19 MB TIF)

Figure S2 Castration reduces the number of genes regulated by infection in males. Genes modulated by *C. burnetii* infection in sterilized or intact male and female mice are represented by a Venn diagram.
Found at: doi:10.1371/journal.pone.0012190.s002 (1.46 MB TIF)

Figure S3 Overview of the tissular components of the biological clock. The heterodimer Clock/Arntl positively regulates transcription of the Per and Cry genes, which in turn negatively regulate Clock and Arntl transcription, creating cyclic expression of these proteins.
Found at: doi:10.1371/journal.pone.0012190.s003 (1.78 MB TIF)

Figure S4 Time course of circadian gene modulation in castrated animals after infection. Castrated male (squares) and ovariectomized female (triangles) mice were infected with *C. burnetii* for one, four and seven days, and gene expression was determined by qRT-PCR and normalized to the GAPDH gene. A–C, Genes involved in circadian rhythm. Horizontal lines indicate the selected fold change cut-off of ±1.75.
Found at: doi:10.1371/journal.pone.0012190.s004 (4.08 MB TIF)

Table S1 Inversed fold changes in males and females. Found at: doi:10.1371/journal.pone.0012190.s005 (0.05 MB PDF)

Table S2 Fold changes assessed by microarray and RT-PCR. Found at: doi:10.1371/journal.pone.0012190.s006 (0.04 MB PDF)

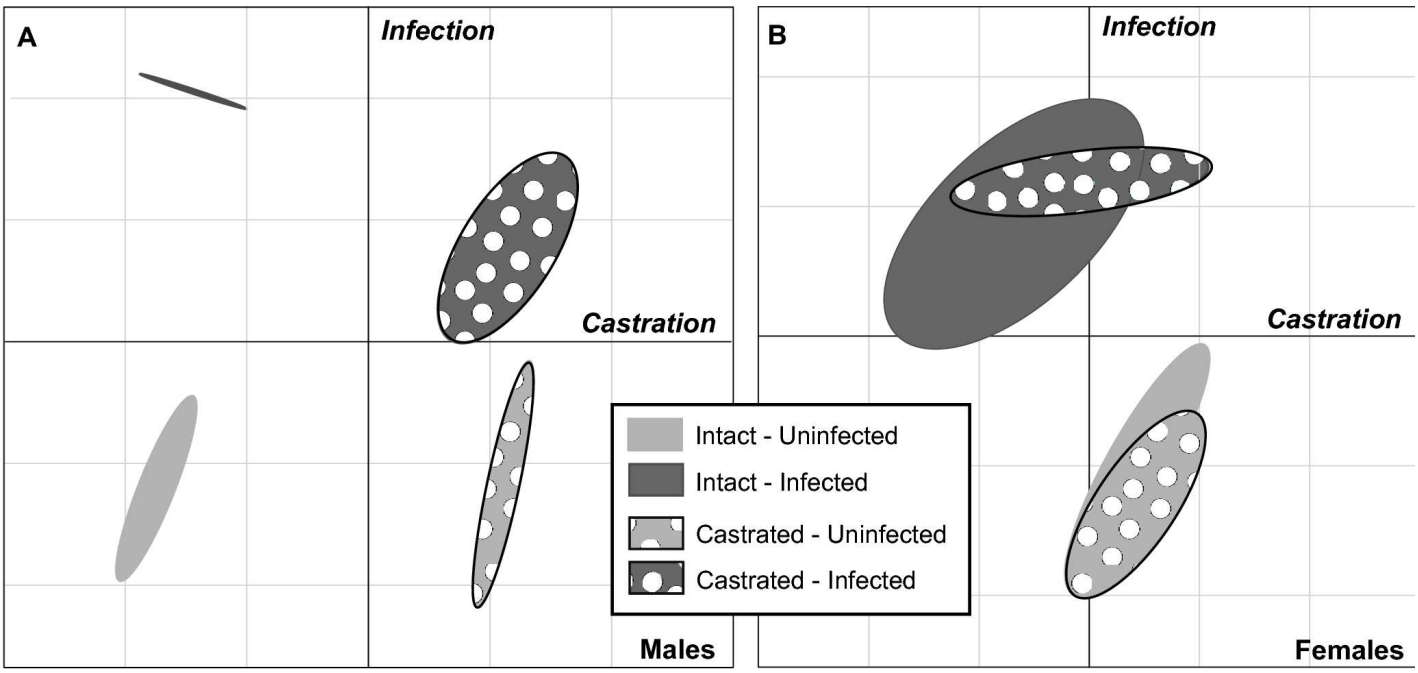
Author Contributions

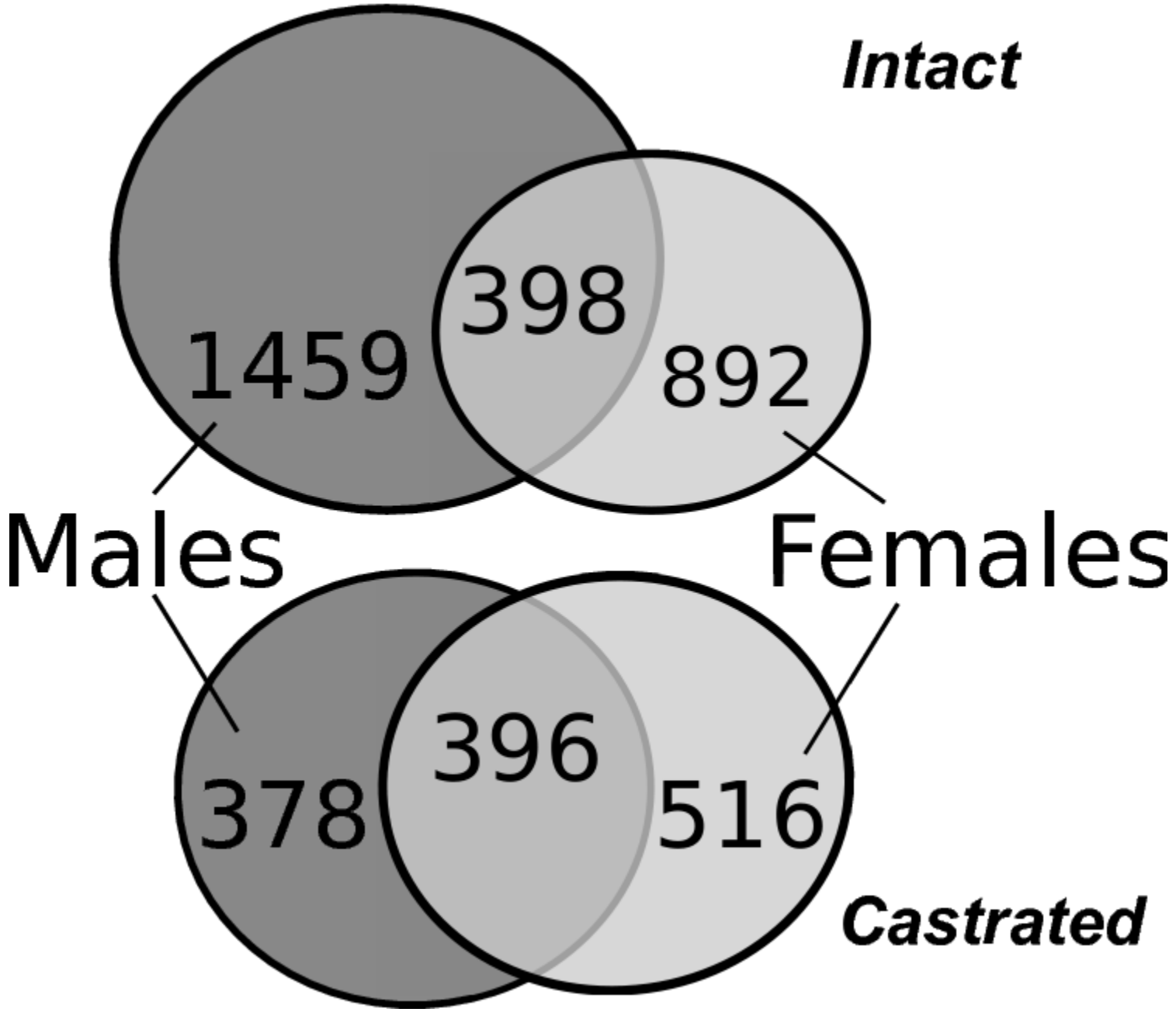
Conceived and designed the experiments: JT DR ML JLM. Performed the experiments: LHB CC. Analyzed the data: JT LHB JLM. Contributed reagents/materials/analysis tools: DR. Wrote the paper: JT CC ML JLM.

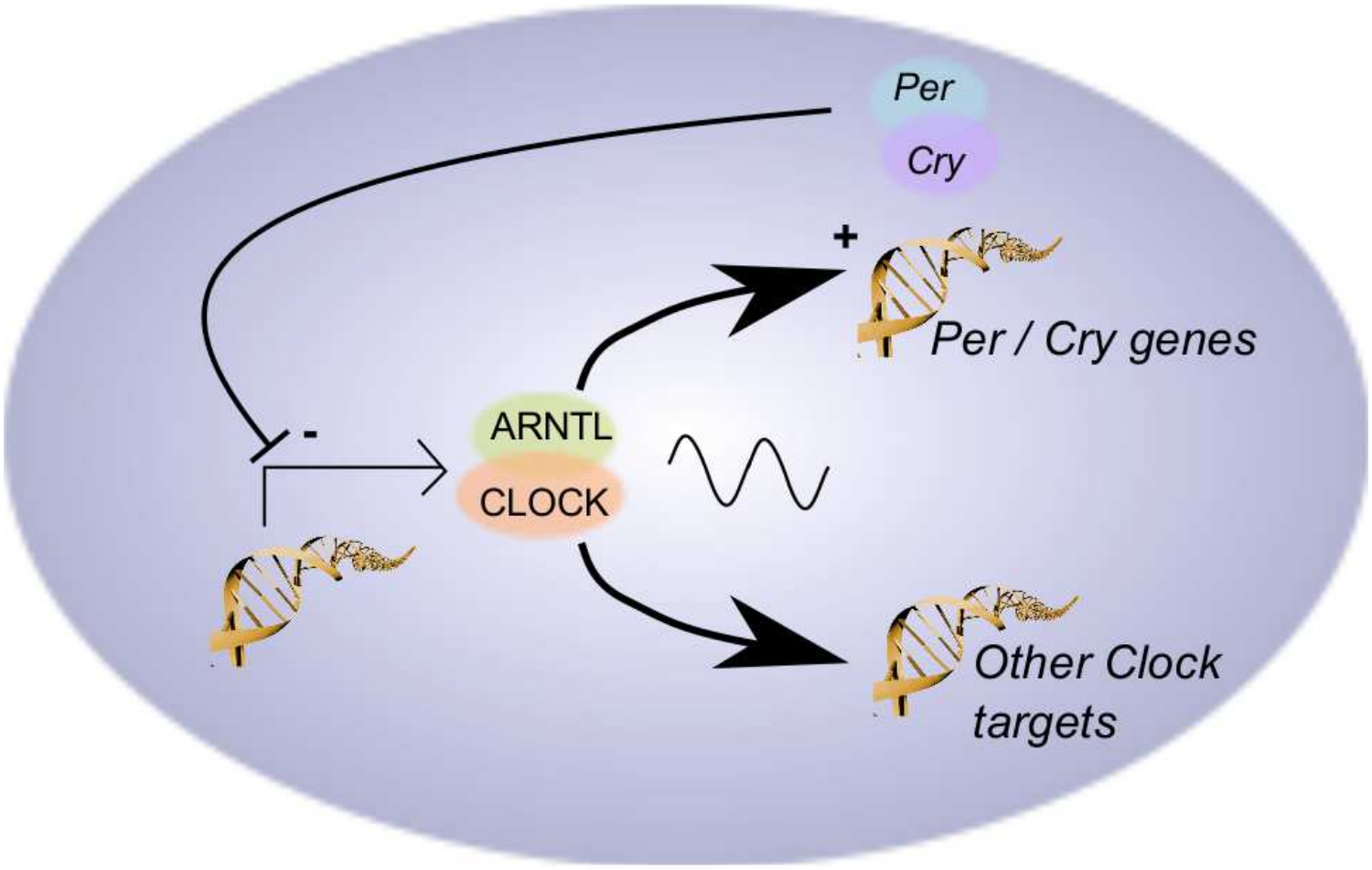
References

1. Theobald S, Tolhurst R, Squire SB (2006) Gender, equity: new approaches for effective management of communicable diseases. Trans R Soc Trop Med Hyg 100: 299–304.
2. Martin GS, Mannino DM, Eaton S, Moss M (2003) The epidemiology of sepsis in the United States from 1979 through 2000. N Engl J Med 348: 1546–1554.
3. Bird MD, Karavitis J, Kovacs EJ (2008) Sex differences and estrogen modulation of the cellular immune response after injury. Cell Immunol 252: 57–67.
4. Holmes CB, Hausler H, Nunn P (1998) A review of sex differences in the epidemiology of tuberculosis. Int J Tuberc Lung Dis 2: 96–104.
5. Jiménez-Corona M, García-García L, DeRiemer K, Ferreyra-Reyes L, Bobadilla-del-Valle M, et al. (2006) Gender differentials of pulmonary tuberculosis transmission and reactivation in an endemic area. Thorax 61: 348–353.
6. Greig JE, Carnie JA, Tallis GF, Ryan NJ, Tan AG, et al. (2004) An outbreak of Legionnaires' disease at the Melbourne Aquarium, April 2000: investigation and case-control studies. Med J Aust 180: 566–572.
7. Hsieh Y, Frink M, Choudhry MA, Bland KI, Chaudry IH (2007) Metabolic modulators following trauma sepsis: sex hormones. Crit Care Med 35: S621–629.
8. Azad AF (2007) Pathogenic rickettsiae as bioterrorism agents. Clin Infect Dis 45: S52–55.
9. Pierce CY, Barr JR, Woolfitt AR, Moura H, Shaw EI, et al. (2007) Strain and phase identification of the U.S. category B agent *Coxiella burnetii* by matrix assisted laser desorption/ionization time-of-flight mass spectrometry and multivariate pattern recognition. Anal Chim Acta 583: 23–31.
10. Raoult D, Marrie T, Mege JL (2005) Natural history and pathophysiology of Q fever. Lancet Infect Dis 5: 219–226.
11. Maurin M, Raoult D (1999) Q fever. Clin Microbiol Rev 12: 518–553.
12. Houpikian P, Raoult D (2005) Blood culture-negative endocarditis in a reference center: etiologic diagnosis of 348 cases. Medicine (Baltimore) 84: 162–173.
13. Leone M, Honstettre A, Lepidi H, Capo C, Bayard F, et al. (2004) Effect of sex on *Coxiella burnetii* infection: protective role of 17beta-estradiol. J Infect Dis 189: 339–345.
14. Penaloza C, Estevez B, Orlanski S, Sikorska M, Walker R, et al. (2009) Sex of the cell dictates its response: differential gene expression and sensitivity to cell death inducing stress in male and female cells. FASEB J 23: 1869–1879.
15. Amador-Noguez D, Zimmerman J, Venable S, Darlington G (2005) Gender-specific alterations in gene expression and loss of liver sexual dimorphism in the long-lived Ames dwarf mice. Biochem Biophys Res Commun 332: 1086–1100.
16. Sopena N, Sabrià M (2005) Multicenter study of hospital-acquired pneumonia in non-ICU patients. Chest 127: 213–219.
17. Yeretssian G, Doiron K, Shao W, Leavitt BR, Hayden MR, et al. (2009) Gender differences in expression of the human caspase-12 long variant determines susceptibility to *Listeria monocytogenes* infection. Proc Natl Acad Sci USA 106: 9016–9020.
18. Basant A, Rege M, Sharma S, Sonawat HM (2010) Alterations in urine, serum and brain metabolomic profiles exhibit sexual dimorphism during malaria disease progression. Malar J 9: 110.
19. Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, et al. (2008) Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. Crit Care Med 36: 296–327.
20. Mokart D, Leone M, Sannini A, Brun JP, Tison A, et al. (2005) Predictive perioperative factors for developing severe sepsis after major surgery. Br J Anaesth 95: 776–781.
21. Moxley G, Posthuma D, Carlson P, Estrada E, Han J, et al. (2002) Sexual dimorphism in innate immunity. Arthritis Rheum 46: 250–258.
22. Marriott I, Bost KL, Huet-Hudson YM (2006) Sexual dimorphism in expression of receptors for bacterial lipopolysaccharides in murine macrophages: a possible mechanism for gender-based differences in endotoxic shock susceptibility. J Reprod Immunol 71: 12–27.
23. van de Beek D, de Gans J, Spanjaard L, Weisfelt M, Reitsma JB, et al. (2004) Clinical features and prognostic factors in adults with bacterial meningitis. N Engl J Med 351: 1849–1859.
24. Hollm-Delgado M, Allard R, Pilon PA (2005) Invasive group A streptococcal infections, clinical manifestations and their predictors, Montreal, 1995–2001. Emerging Infect Dis 11: 77–82.
25. Treister NS, Richards SM, Suzuki T, Jensen RV, Sullivan DA (2005) Influence of androgens on gene expression in the BALB/c mouse submandibular gland. J Dent Res 84: 1187–1192.
26. Davidoff AM, Ng CYC, Zhou J, Spence Y, Nathwani AC (2003) Sex significantly influences transduction of murine liver by recombinant adeno-associated viral vectors through an androgen-dependent pathway. Blood 102: 480–488.
27. Ceribelli A, Pino MS, Cecere FL (2007) Gender differences: implications for clinical trials and practice. J Thorac Oncol 2: S15–18.
28. Paretsky D, Downs CM, Salmon CW (1964) Some biochemical changes in the guinea pig during infection with *Coxiella burnetii*. J Bacteriol 88: 137–142.
29. Howe D, Heinzen RA (2006) *Coxiella burnetii* inhabits a cholesterol-rich vacuole and influences cellular cholesterol metabolism. Cell Microbiol 8: 496–507.
30. Esemuede N, Lee T, Pierre-Paul D, Sumpio BE, Gahtan V (2004) The role of thrombospondin-1 in human disease. J Surg Res 122: 135–142.
31. He Y, Li H, Zhang J, Hsu C, Lin E, et al. (2004) The extracellular matrix protein mindin is a pattern-recognition molecule for microbial pathogens. Nat Immunol 5: 88–97.
32. Capo C, Lindberg FP, Meconi S, Zaffran Y, Tardei G, et al. (1999) Subversion of monocyte functions by *Coxiella burnetii*: impairment of the cross-talk between αvβ3 integrin and CR3. J Immunol 163: 6078–6085.
33. Froy O (2005) Regulation of mammalian defensin expression by Toll-like receptor-dependent and independent signalling pathways. Cell Microbiol 7: 1387–1397.
34. Kaiser V, Diamond G (2000) Expression of mammalian defensin genes. J Leukoc Biol 68: 779–784.
35. Nemeth E, Ganz T (2009) The role of hepcidin in iron metabolism. Acta Haematol 122: 78–86.
36. Briggs HL, Pul N, Seshadri R, Wilson MJ, Tersteeg C, et al. (2008) Limited role for iron regulation in *Coxiella burnetii* pathogenesis. Infect Immun 76: 2189–2201.
37. Courselaud B, Troadec M, Fruchon S, Ilyin G, Borot N, et al. (2004) Strain and gender modulate hepatic hepcidin 1 and 2 mRNA expression in mice. Blood Cells Mol Dis 32: 283–289.
38. Krijt J, Cmejla R, Sýkora V, Vokurka M, Vyoral D, et al. (2004) Different expression pattern of hepcidin genes in the liver and pancreas of C57BL/6N and DBA/2N mice. J Hepatol 40: 891–896.
39. Meghari S, Capo C, Raoult D, Mege JL (2006) Deficient transendothelial migration of leukocytes in Q fever: the role played by interleukin-10. J Infect Dis 194: 365–369.
40. Mege JL, Meghari S, Honstettre A, Capo C, Raoult D (2006) The two faces of interleukin 10 in human infectious diseases. Lancet Infect Dis 6: 557–569.
41. Ghigo E, Honstettre A, Capo C, Gorvel J, Raoult D, et al. (2004) Link between impaired maturation of phagosomes and defective *Coxiella burnetii* killing in patients with chronic Q fever. J Infect Dis 190: 1767–1772.
42. Dellacasagrande J, Ghigo E, Raoult D, Capo C, Mege JL (2002) IFN-gamma-induced apoptosis and microbicidal activity in monocytes harboring the intracellular bacterium *Coxiella burnetii* require membrane TNF and homotypic cell adherence. J Immunol 169: 6309–6315.
43. Kahlke V, Angele MK, Ayala A, Schwacha MG, Cioffi WG, et al. (2000) Immune dysfunction following trauma-haemorrhage: influence of gender and age. Cytokine 12: 69–77.
44. Miller BH, McDearmon EL, Panda S, Hayes KR, Zhang J, et al. (2007) Circadian and CLOCK-controlled regulation of the mouse transcriptome and cell proliferation. Proc Natl Acad Sci USA 104: 3342–3347.
45. Haimovich B, Calvano J, Haimovich AD, Calvano SE, Coyle SM, et al. (2010) In vivo endotoxin synchronizes and suppresses clock gene expression in human peripheral blood leukocytes. Crit Care Med 38: 751–758.
46. Chappell PE (2005) Clocks and the black box: circadian influences on gonadotropin-releasing hormone secretion. J Neuroendocrinol 17: 119–130.
47. Cai W, Rambaud J, Teboul M, Masse I, Benoit G, et al. (2008) Expression levels of estrogen receptor beta are modulated by components of the molecular clock. Mol Cell Biol 28: 784–793.
48. Gery S, Virk RK, Chumakov K, Yu A, Koeffler HP (2007) The clock gene Per2 links the circadian system to the estrogen receptor. Oncogene 26: 7916–7920.
49. Iwahana E, Karatsoreos I, Shibata S, Silver R (2008) Gonadectomy reveals sex differences in circadian rhythms and suprachiasmatic nucleus androgen receptors in mice. Horm Behav 53: 422–430.
50. Karatsoreos IN, Wang A, Sasanian J, Silver R (2007) A role for androgens in regulating circadian behavior and the suprachiasmatic nucleus. Endocrinology 148: 5487–5495.
51. Habbal OA, Al-Jabri AA (2009) Circadian rhythm and the immune response: a review. Int Rev Immunol 28: 93–108.
52. Lee J, Edery I (2008) Circadian regulation in the ability of *Drosophila* to combat pathogenic infections. Curr Biol 18: 195–199.
53. Arjona A, Sarkar DK (2006) The circadian gene mPer2 regulates the daily rhythm of IFN-gamma. J Interferon Cytokine Res 26: 645–649.

54. Arjona A, Sarkar DK (2005) Circadian oscillations of clock genes, cytolytic factors, and cytokines in rat NK cells. J Immunol 174: 7618–7624.
55. Nader N, Chrousos GP, Kino T (2009) Circadian rhythm transcription factor CLOCK regulates the transcriptional activity of the glucocorticoid receptor by acetylating its hinge region lysine cluster: potential physiological implications. FASEB J 23: 1572–1583.
56. Farbrother P, Wagner C, Na J, Tunggal B, Morio T, et al. (2006) *Dictyostelium* transcriptional host cell response upon infection with *Legionella*. Cell Microbiol 8: 438–456.
57. Bastonero S, Le Priol Y, Armand M, Bernard CS, Reynaud-Gaubert M, et al. (2009) New microbicidal functions of tracheal glands: defective anti-infectious response to *Pseudomonas aeruginosa* in cystic fibrosis. PLoS ONE 4: e5357.
58. Schmittgen TD, Zakrajsek BA, Mills AG, Gorn V, Singer MJ, et al. (2000) Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. Anal Biochem 285: 194–204.
59. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19: 185–93.
60. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98: 5116–21.
61. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.
62. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, et al. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4: P3.
63. Culhane AC, Thioulouse J, Perrière G, Higgins DG (2005) MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics 21: 2789–2790.
64. Barrett T, Edgar R (2006) Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Meth Enzymol 411: 352–369.
65. Lopez F, Textoris J, Bergon A, Didier G, Remy E, et al. (2008) TranscriptomeBrowser: a powerful and flexible toolbox to explore productively the transcriptional landscape of the Gene Expression Omnibus database. PLoS ONE 3: e4001.


A
Infection
Castration
Males
B
Infection
Castration
Females
Intact - Uninfected
Intact - Infected
Castrated - Uninfected
Castrated - Infected


Intact
1459
398
892
Males
Females
396
378
516
Castrated


Per
Cry
+
Per / Cry genes
-
ARNTL
CLOCK
Other Clock
targets

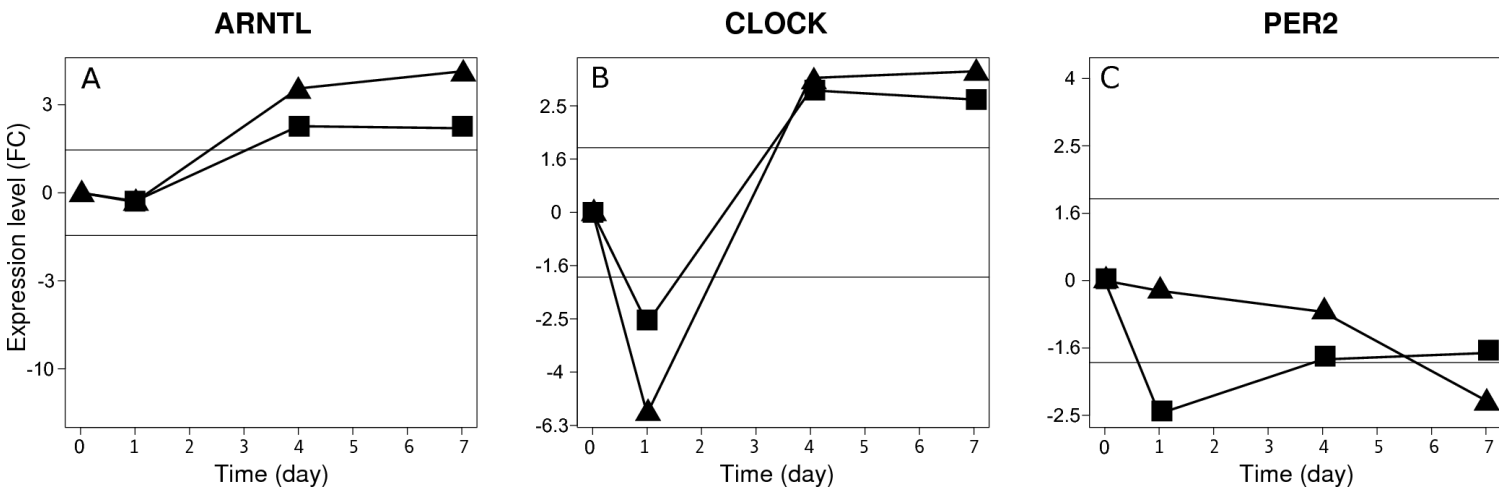

ARNTL
CLOCK
PER2
A
B
C
Expression level (FC)
3
0
-3
-10
2.5
1.6
0
-1.6
-2.5
-4
-6.3
4
2.5
1.6
0
-1.6
-2.5
0 1 2 3 4 5 6 7
Time (day)

Table S1. Inversed fold changes in males and females

down-modulated in males up-regulated in females			up-regulated in males down-modulated in females		
BB505010	- 12.5	2.6	D930050A07Rik	1.8	- 2.0
Gm534	- 6.7	1.9	PPP1R3C	1.8	- 2.2
AK007854	- 5.6	1.9	CHKA	1.9	- 2.6
LHX8	- 3.1	1.9	BAIAP2	1.9	- 2.7
CYP3A44	- 2.4	2.7	MLL5	1.9	- 2.0
AACS	- 2.3	2.8	MYBPC2	1.9	- 2.2
RGS16	- 2.2	11.5	FST	1.9	- 1.8
AK051762	- 2.1	1.8	CSPG5	2.0	- 2.7
OPRM1	- 2.0	1.9	BC057022	2.1	- 1.9
6030468B19Rik	- 2.0	1.9	AK040776	2.3	- 1.9
Gm7231	- 1.9	3.4	LOXL4	2.4	- 4.3
OLFR1321	- 1.9	2.3	EMID2	2.9	- 1.8
BAI3	- 1.8	1.9	COG4	3.2	- 1.8
Gm10001	- 1.8	2.0	OLFR914	3.8	- 3.0

Table S2. Fold changes assessed by microarray and RT-PCR

	Fold change			
	microarray		**RT-PCR**	
	males	females	males	females
IL6	1.3	2.2	1.3	1.2
IFNG	- 1.1	- 1.1	- 1.7	1.0
IL10	1.7	1.2	2.1	1.2
ARNTL	1.4	- 11.1	1.0	-18
CLOCK	1.1	- 2.6	1.0	- 4.7
PER2	1.2	3.91	1.1	4.0

Gene Expression Signature-Based Screening Identifies new Broadly Effective Influenza A Antivirals

Laurence Josset, **Julien Textoris**, Béatrice Loriod, Olivier Ferraris, Vincent Moules, Bruno Lina, Catherine N'Guyen, Jean-Jacques Diaz, Manuel Rosa-Calatrava

PLoS ONE 5(10): e13169

Les anti-viraux classiques ciblent des protéines virales et sont donc exposés à l'émergence de résistances. Pour contourner ce problème, l'idée de s'attaquer aux protéines de l'hôte a été proposée comme stratégie alternative. Notre hypothèse est que cette approche permettrait d'identifier de nouveaux antiviraux à large spectre. Le virus *Influenza A* a été utilisé comme modèle en raison de sa diversité antigénique. De plus, comme l'a souligné la récente pandémie H1N1, il est nécessaire de développer des stratégies thérapeutiques permettant de faire face à l'émergence d'un nouveau variant. Nous avons donc proposé l'identification d'une signature transcriptionnelle de réponse à l'infection par différentes souches de virus de la grippe. L'idée est que cette réponse commune permettrait d'identifier de potentielles cibles cellulaires permettant le développement d'antiviraux à large spectre.

Nous avons donc analysé les modulations du transcriptome induites lors de l'infection par cinq souches humaines ou aviaires du virus *Influenza A*. Cette approche a conduit à l'identification d'une signature transcriptionnelle de 300 gènes dont l'expression est modulée entre les cellules infectées et non infectées. A partir de cette signature, une liste des vingts gènes présentant la plus forte modulation a été utilisée pour interroger la base *ConnectivityMap.* Cette base de données regroupe des signatures transcriptionnelles induites par diverses molécules. Les molécules induisant un profil d'expression inverse à celui obtenu après l'infection ont été choisies comme potentiels candidats antiviraux. Notre hypothèse est que ces molécules entraîneraient un environnement défavorable à la réplication virale. Huit antiviraux potentiels dont la ribavirine ont été sélectionnés et leur effet sur la réplication des cinq souches a été évalué *in-vitro*. Six de ces molécules inhibent la réplication virale. De plus, la réplication du virus H1N1 responsable de la récente pandémie était inhibée par cinq de ces huits molécules, alors que cette souche n'avait pas été utilisée pour définir la signature transcriptionnelle.

Ce résultat démontre que notre stratégie permet d'identifier des antiviraux à large spectre, actif sur des souches virales émergentes. Les gènes de la signature transcriptionnelle identifiée codent probablement pour des protéines cellulaires impliquées dans le cycle viral. Il s'agit de la première étude démontrant le potentiel de l'analyse du transcriptome pour identifier de nouvelles cibles thérapeutiques. Cette approche accélère considérablement le développement de nouvelles molécules et pourrait être envisagée pour

d'autres modèles infectieux.

Ce travail a donné lieu au dépôt d'un brevet national (09-58810 en date du 9 décembre 2009), américain (62/267,997, en date du 9 décembre 2009) et international portant le numéro PCT/EP2010/069023 (Office Européen des Brevets).

OPEN ACCESS Freely available online

PLoS one

Gene Expression Signature-Based Screening Identifies New Broadly Effective Influenza A Antivirals

Laurence Josset[1,2]*, Julien Textoris[3,4,5], Béatrice Loriod[3], Olivier Ferraris[1], Vincent Moules[1], Bruno Lina[1,2], Catherine N'Guyen[3], Jean-Jacques Diaz[4], Manuel Rosa-Calatrava[1]*

1 Centre National de la Recherche Scientifique (CNRS) FRE 3011 Virologie et Pathologie Humaine, Université Lyon 1, Lyon, France, **2** Laboratoire de Virologie Centre de Biologie et de Pathologie Est, Hospices Civils de Lyon, Lyon, France, **3** Institut National de la Santé et de la Recherche Médicale (INSERM) U928 Technologies Avancées pour le Génome et la Clinique, Université de la Méditerranée, Marseille, France, **4** Centre National de la Recherche Scientifique (CNRS) UMR 5534, Centre Léon Bérard, Centre de Génétique Moléculaire et Cellulaire, Université Lyon 1, Lyon, France, **5** Service d'anesthésie et de réanimation Hôpital Nord, Assistance Publique - Hôpitaux de Marseille, Marseille, France

Abstract

Classical antiviral therapies target viral proteins and are consequently subject to resistance. To counteract this limitation, alternative strategies have been developed that target cellular factors. We hypothesized that such an approach could also be useful to identify broad-spectrum antivirals. The influenza A virus was used as a model for its viral diversity and because of the need to develop therapies against unpredictable viruses as recently underlined by the H1N1 pandemic. We proposed to identify a gene-expression signature associated with infection by different influenza A virus subtypes which would allow the identification of potential antiviral drugs with a broad anti-influenza spectrum of activity. We analyzed the cellular gene expression response to infection with five different human and avian influenza A virus strains and identified 300 genes as differentially expressed between infected and non-infected samples. The most 20 dysregulated genes were used to screen the connectivity map, a database of drug-associated gene expression profiles. Candidate antivirals were then identified by their inverse correlation to the query signature. We hypothesized that such molecules would induce an unfavorable cellular environment for influenza virus replication. Eight potential antivirals including ribavirin were identified and their effects were tested in vitro on five influenza A strains. Six of the molecules inhibited influenza viral growth. The new pandemic H1N1 virus, which was not used to define the gene expression signature of infection, was inhibited by five out of the eight identified molecules, demonstrating that this strategy could contribute to identifying new broad anti-influenza agents acting on cellular gene expression. The identified infection signature genes, the expression of which are modified upon infection, could encode cellular proteins involved in the viral life cycle. This is the first study showing that gene expression-based screening can be used to identify antivirals. Such an approach could accelerate drug discovery and be extended to other pathogens.

Citation: Josset L, Textoris J, Loriod B, Ferraris O, Moules V, et al. (2010) Gene Expression Signature-Based Screening Identifies New Broadly Effective Influenza A Antivirals. PLoS ONE 5(10): e13169. doi:10.1371/journal.pone.0013169

Editor: Man-Seong Park, Hallym University, Republic of Korea

Received June 2, 2010; **Accepted** September 9, 2010; **Published** October 4, 2010

Funding: This work was supported by CNRS (Centre National de la Recherche Scientifique), the Hospices Civils de Lyon, the European Vigilance Network for the Management of Antiviral Drug Resistance (VIRGIL), and by a grant of the "Programme de Recherche A(H1N1)" co-ordinated by the Institut de Microbiologie et Maladies Infectieuses (INSERM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: laurence.josset@chu-lyon.fr (LJ); manuel.rosa-calatrava@univ-lyon1.fr (MRC)

Introduction

Antiviral drug development is currently based on two approaches: i) the conventional approach of inhibiting the activity of a viral enzyme which often leads to the emergence of drug resistant viruses due to viral genomic variability and ii) the more recent approach of targeting cellular factors that are required for viral replication. Indeed, coding for a limited number of proteins, viruses hijack the cellular machinery and rely on many host proteins for their replication. The major recognized advantage of targeting a host factor is therefore to limit the development of resistance as the virus cannot replace a missing cellular protein [1]. Such an approach has been used in antiretroviral therapy with the development of a CCR5 antagonist showing promise as an anti-HIV drug [2]. We have also demonstrated that this strategy is efficient at inhibiting the replication of herpes viruses resistant to conventional antivirals [3]. In influenza research, the effective in vitro and in vivo inhibition of two different cellular pathways without inducing resistance has been reported, and both are currently undergoing preclinical trials (recently reviewed in [4]).

Targeting cellular proteins may provide another crucial advantage: if a cellular pathway is critical to the viral cycle, agents that target such a pathway should represent potential broad-spectrum antivirals. The influenza virus represents a constant threat to public health due to the emergence of new viral strains and is therefore an ideal model on which to test this hypothesis.

Belonging to the orthomyxoviridae family, influenza viruses have genomes composed of single-stranded RNA and are classified into three types: A, B and C according to their internal protein sequences [5]. The influenza A viruses are further subtyped based on the antigenicity of the two envelope glycoproteins hemagglutinin (HA) [H1 to H16] and neuraminidase (NA) [N1 to N9]. All influenza A subtypes are endemic in aquatic birds but only two,

H1N1 and H3N2, are presently circulating among humans. Since the influenza genome is segmented, two different viral strains infecting the same cell are able to reassort their genomic segments. Variability can also be due to the low fidelity of the viral RNA polymerase, which causes yearly epidemics owing to an antigenic drift in glycoproteins. Novel pathogenic strains of the influenza virus have also emerged with antigenically different HA and/or NA and have caused three pandemics in the 20th century: the Spanish influenza (H1N1) in 1918, responsible for approximately 50 million deaths; the Asian influenza (H2N2) in 1957 during which about 2–4 million people died; and the Hong Kong influenza (H3N2) in 1968 responsible for 1–2 million deaths [6].

Considering this pandemic potential and with up to 500,000 annual deaths worldwide during usual winter outbreaks, influenza A viruses represent a major public health concern [7]. Prevention relies on vaccination which has several major limitations including the lag time for vaccine preparation and the low vaccination coverage rate. Once a patient becomes infected, the current etiologic treatment of flu relies on M2 channel blockers or NA inhibitors [8]. However, these existing therapies are inappropriate for use in cases of severe infection and may be limited due to the risk of rapid emergence of drug resistant viruses. Thus there is an obvious need to complement existing therapies with new anti-influenza drugs.

To search for new antivirals, we hypothesized that common viral effects on cell metabolism should occur after infection with different avian and human influenza viruses and that this pattern should lead to the identification of drugs effective on all influenza A viruses potentially. We first sought to identify a common gene expression signature following the infection with different human and avian influenza A viruses. While several microarray analyses have already compared the pandemic 1918 H1N1 virus [9,10] or some H5N1 strain [11,12] to other less pathogenic strains, our study is the first to demonstrate that a global influenza-induced gene-expression signature can be defined. This proof-of-concept study was conducted on a home-made nylon array using a human pulmonary epithelial cell line infected by five influenza A virus subtypes (H1N1, H3N2, H5N1, H5N2 and H7N1). Using this signature, we determined if molecules disturbing this pattern of infection would have a broad-influenza antiviral effect. By consulting the Connectivity Map, a database of drug-associated gene expression profiles [13,14], we identified molecules that induced gene expression changes after cell treatment that were mainly opposite to those induced by infection. These molecules were tested in vitro for their effect on the five different viruses. To confirm our methodology, we took the opportunity of using the new emerging pandemic H1N1 virus as a model to test the effect of these molecules on a new unknown virus.

Materials and Methods

1 Cell lines and viruses

Cells of the human lung epithelial cell line A549 were grown as monolayers in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 100 U of penicillin/mL, and 100 µg of streptomycin sulfate/mL at 37°C.

Influenza viruses A/New Caledonia/20/99 (H1N1), A/Moscow/10/99 (H3N2), A/Lyon/969/09 (H1N1 SOIV), A/Turkey/582/2006 (H5N1), A/Finch/England/2051/94 (H5N2), and A/Chicken/Italy/2076/99 (H7N1) were produced in MDCK cells in EMEM supplemented with 2 mM L-glutamine, 100U of penicillin/mL, 100 µg of streptomycin sulfate/mL and 1 µg of trypsin/mL. Viruses were titrated to determine tissue culture infection dose 50% (TCID50) in MDCK cells as described in our previous study [15].

For the microarray analysis, A549 cells were infected for 24 h at 37°C with influenza viruses at a multiplicity of infection (moi) of 1 in DMEM supplemented with 2 mM L-glutamine, 100 U of penicillin/mL, 100 µg of streptomycin sulfate/mL and 0.5 µg of trypsin/mL (infection medium). This moi was chosen to ensure that 100% of the cells were infected 24 h postinfection. The microarray experiments were performed in five independent replicates.

For kinetics on A549 cells, confluent cells were infected with influenza viruses at a moi of 0.1 or 2 for one hour under a minimal volume of infection medium at 37°C. The cells were then overlaid with fresh infection medium and incubated at 37°C. Samples of supernatants were collected at defined time points and stored at −80°C until end point titration assays (TCID50) in MDCK cells.

2 RNA preparation and hybridization to the gene chip

Total RNA was extracted from cell pellets using an RNeasy Mini Kit (Qiagen,Valencia,CA) for the BSL2 viruses. For H5N1 infections, total RNA was extracted with Trizol LS (Invitrogen). mRNAs were labeled with ^{33}P for the reverse transcription using the Superscript III RT (Invitrogen), (α33P)dCTP and an oligodT25. Generated cDNAs were hybridized on home-made Nylon microarrays (HuSG9k) containing 9216 spotted IMAGE human cDNA clones, representing 8682 genes and 434 control clones [16]. Further details on the HuSG9k microarray are available on the TAGC website (http://tagc.univ-mrs.fr/). All membranes used in this study belonged to the same batch. After hybridization and exposure on Micro Imager, arrays were scanned in a Fuji BAS 5000 machine and hybridization signals quantified using the BZ Scan Software [17]. Primary data, in accordance with the proposed MIAME standards, are accessible through GEO Series accession number GSE22319 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE22319).

3 Data normalization and analysis

Data files were loaded and analyzed with R (v2.9.2) and Bioconductor (v2.4.1) [18], using the NylonArray library developed by the TAGC to support BZScan2 files (library available upon request). Raw data were normalized by quantile normalization. Supervised analysis (supervised methods aim at finding a set of genes whose expression profiles best correlate with a known phenotype) between groups *Infected* and *Mock* samples was conducted using the Significance Analysis of Microarray algorithm (SAM) [19], using the siggenes library (v1.18.0) [20]. All statistical analyses involved corrections for multiple comparisons (Benjamini and Hochberg) [21]. Agglomerative hierarchical clustering was performed by the pairwise average-linkage method using the Pearson correlation distance (Cluster 3.0, Eisen, Stanford University).

4 Quantitative real-time RT-PCR validation

To validate the microarray results with real-time RT-PCR assay, another set of A549 cells were infected with influenza viruses at a moi of 1 and total cell RNA was extracted at 24 hpi with Trizol LS (Invitrogen). Five hundred ng of total RNA were reverse transcribed using oligo(dT)18 and RevertAid M-MuLV (Fermentas) according to the manufacturer's instructions. One µL of cDNA was then amplified and analyzed in the 7500 Real Time PCR System (Applied Biosystems) using the Platinum(R) SYBR(R) Green qPCR SuperMix-UDG kit (Invitrogen) according to the manufacturer's instructions. Six genes were chosen according to their level of expression (Fold change in $\log_2 > 2$ or < -2) and the availability of primers for the quantitative PCR (Table S1). Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) mRNA was

used as an internal control. The reaction mix contained a total volume of 20 µL and the thermal cycling consisted of UDG incubation at 50°C for 2 min, 40 cycles of 95°C for 15 s and 60°C for 33 s for amplification. All data were normalized to the internal standard *GAPDH* mRNA. For each single-well amplification reaction, a threshold cycle (Ct) was observed in the exponential phase of amplification. Relative changes in gene expression were determined using the $2^{\Delta\Delta Ct}$ method as previously described [22] and reported as the n-fold difference relative to a control cDNA (mock cells) prepared in parallel with the experimental cDNAs (infected cells). Statistical significance was calculated using Welch's two sample t-test between mock and infected samples using R software.

5 *In silico* experiment: query the Connectivity Map with the infection signature

To select potential antivirals, an unbiased in silico search for molecules that reverse the infection signature identified in the present study was performed using the publicly available Connectivity Map database (build 02) [13]. The Connectivity Map (also known as CMAP) is a collection of genome-wide transcriptional data from cultured human cells treated with different kinds of molecules. The 20 most differentially expressed genes in the infection state (Fold Change in $\log_2 > 2$ or < -2) were selected from the initial 300 gene set identified by SAM. These were then mapped to the U133A probe sets in order to query the Connectivity Map database. In total, 28 U133A probe sets mapped to the selected genes from this study. The connectivity scores and p-values were obtained using the CMAP algorithm [13].

6 Molecules

2-aminobenzenesulfonamide (Sigma), calcium folinate (Sigma), harmol hydrochloride (MP Biomedical), merbromine (Sigma), midodrine (Sigma) and ribavirin (Valeant Pharmaceuticals) were dissolved in sterile water to a stock concentration of 5 g/L, 5 g/L, 4 g/L, 3.4 g/L, 5 g/L and 10 mM respectively. Rilmenidine (Sigma) was dissolved in dimethylsulfoxide (DMSO) to a stock concentration of 13 g/L and brinzolamide was in suspension at 10 g/L in the collyrium AZOPT.

Sulfameter (Sigma), pyrvinium (Sigma), moxalactam (Sigma) and methylbenzethoniumchloride (Sigma) were dissolved in sterile water to a stock concentration of 50 g/L. Alvespimycin (Sigma) was dissolved in sterile water to a concentration of 0.03 g/L. Sulodictil (Sigma) and DL-Thiorphan (Sigma) were dissolved in DMSO to a concentration of 50 g/L.

7 Viability assays

Cell viability was measured by the neutral red assay, an indicator of cytotoxicity used in cultures of different cell lines [23] with the same sensitivity as the MTT assay [24,25]. The neutral red assay is based on the initial protocol described by Borenfreund and Puerner (1984) and determines the accumulation of the neutral red dye in the lysosomes of viable, uninjured cells. Cells were seeded into 96-well plates and treated with molecules or solvent. 72 h after treatment, cells were incubated for 3 h with neutral red dye (100 µg/ml) dissolved in serum free medium (DMEM). Cells were then washed with phosphate buffered saline (PBS) and fixed in a formol/calcium mix (40%/10%) for 1 min before being lysed with EtOH/AcCOOH, (50%/1%) followed by gentle shaking for 15 min until complete dissolution was achieved. Absorbance at 550 nm was measured using a microplate spectrophotometer system (Microplate Reader 2001, BioWhittaker) and results were presented as a ratio of control values.

8 Neuraminidase assay

Standard fluorometric endpoint assays used to monitor NA activity was recently shown to be suitable to quantify influenza virus in a high-throughput screening test [26]. Briefly, cell supernatants (25 µl) were transferred to a black 96-well plate and 75 µl of 2′-(4-methylumbelliferyl)-alpha-*N*-acetylneuraminic acid (MUNANA, Sigma Chemical Co.) to a final concentration of 50 µM were added. After incubation of the plate at 37°C for 1 hr, 150 µl stop solution (0.05 M glycine, pH 10.4) was added to each well and the fluorescence read on a FluoStar Opima (BMG Labtech) with excitation and emission filters of 355 nm and 460 nm respectively. Relative fluorescence units (RFU) were corrected by subtracting specific blanks, ie medium with or without molecules.

For the NA activity test on L3 viruses (H5N1), viruses were inactivated as previously described [27]. Cell supernatants were mixed with freshly prepared Triton X-100 to a final concentration of 1% (vol/vol) Triton X-100 and incubated for 1 h at room temperature. The inactivated supernatants were then transported out of the BSL3 to the BSL2 laboratory and used for NA assays as described above.

Potential interference of test molecules on the NA enzymatic activity was tested by incubating the A/Moscow/10/99 (H3N2) viral stock diluted in DMEM ($10^{7.8}$ $TCID_{50}$/mL final) with increasing concentrations of the test molecule (or DMEM for control) for 0.5 h at room temperature. Specific blanks were measured for each molecule. 25 µL were used for the NA test as described above and results were expressed as a ratio of corrected RFU of the sample to RFU of controls. Two independent experiments were performed in duplicate.

9 Viral growth assays in the presence of the molecules

For the viral growth assays in the presence of the molecules, A549 cells were seeded into 96-well plates at 0.15×10^5 cells per well and cultured for 3 days to 100% confluence. Cells were then washed with DMEM and incubated with various concentrations of the different molecules diluted in infection medium (DMEM supplemented with 2 mM L-glutamine, 100 U/mL penicillin, 100 µg/mL streptomycin, 20 mM HEPES and 0.5 µg/mL trypsin). Six hours after treatment, cells were infected with influenza viruses at a moi of 2 or 0.2 by adding 25 µL per well of virus diluted in infection medium. Infection was allowed to proceed for 65 h at 37°C, 5% CO2 after which 25 µL of supernatant were collected for the NA activity test. Results are expressed as a ratio of corrected RFU of the sample to RFU of control (incubation with infection medium without test molecule). To check for cytotoxicity, viability assays were performed in parallel to each viral growth assay.

10 Test of infection efficiency after cell or virus pre-incubation with the molecules

A549 cells were seeded into 96-well plates at 0.15×10^5 cells per well and cultured for 3 days to 100% confluence. For the 'Cell Preincubation' test, cells were washed with DMEM and incubated with various concentrations of the different molecules diluted in 200 µL per well of infection medium for 14 h. After two washings with DMEM, cells were infected with influenza A/Moscow/10/99 (H3N2) virus at a moi of 7 during 15 min and washed twice with infection medium. Infection was allowed to proceed for 5 h at 37°C. For the 'Virus Preincubation' assay, the molecules were diluted in infection medium and A/Moscow/10/99 (H3N2) viral stock ($10^{8.8}$ $TCID_{50}$/mL) was treated with increasing concentrations of the molecules for 14 h. Cells were then washed with

DMEM and incubated for 15 min with the virus and molecule mix diluted 12 times. Infection was allowed to proceed for 5 h at 37°C. In both assays, the number of infected cells was estimated with a NA test. Cells were washed with PBS and lysed by shaking for 1 h with 25 µL per well of Triton 1X. The cell lysis extracts were used for a neuraminidase test as described above. Results were expressed as a ratio of corrected RFU of sample to RFU of control (solvent treated infections). Statistical significance was calculated in comparison to results for control cells using two tailed Welch t test.

11 EC_{50} and CC_{50} calculations

Viability and antiviral data were analyzed using the following three-parameter non linear logistic regression (l3) function [28]

$$y = D + \frac{D}{1 + \left(\frac{x}{E}\right)^{B}}$$

were y is the response, D is the upper limit (response when the dose x is 'infinite'), E is denoted EC_{50} or CC_{50} and is the dose producing a response half-way between the upper limit and lower limit (0), and B is the relative slope around E. This model is the shortened form of the four parameter logistic function where the lower limit is fixed to 0. Results were obtained by fitting the l3 function using the package drc [29] in the R Statistical Language (version2.7.1). Parameters of the l3 model were estimated and fitted curves were plotted only if the data set contained one response $<D/2$.

Results

1 Global transcriptional signature of influenza A infection

To characterize the global cellular gene-expression response to influenza A infection, human pulmonary epithelium A549 cells were infected with human A/New Caledonia/20/99 (H1N1) and A/Moscow/10/99 (H3N2) and avian A/Turkey/582/2006 (H5N1), A/Finch/England/2051/94 (H5N2), and A/Chicken/Italy/2076/99 (H7N1) influenza viral strains. These viruses are herein referred to as H1N1, H3N2, H5N1, H5N2 and H7N1. A549 cells express both sialic acid α2,6- and α2,3-galactose receptors [30,31] and were shown to be infected by human, avian and swine influenza viruses [32,33]. Infections were performed at 37°C, a temperature at which both human and avian influenza viruses efficiently infect cell cultures [34] and at a moi of 0.1. In these conditions, there was evidence of productive viral replication of all viruses but with some kinetic and yield differences between viruses, as determined by infectious titers ($TCID_{50}$) of supernatants of influenza virus infected A549 cells (Figure 1). The H5N1 virus titers peaked higher and earlier (24 hpi) compared to other viruses titers. Avian H7N1 and H5N2 viruses replicated with correct efficiencies, similar to the human H3N2 virus. In contrast, the human H1N1 virus strain replicated slower (titers peaked at 65 hpi) and grew to lower titers than other viruses (p-value <0.05 at 24 hpi).

To determine the host gene-response to infection, total cellular RNA was extracted at 24 hpi and submitted to reverse transcription in the presence of ^{33}P. Each condition was performed in 5 independent replicates. All labeled cDNAs provided a good radioactive intensity and were hybridized onto home-made nylon microarrays containing 8782 IMAGE cDNA clones. All hybridizations were of good quality according to signals within acceptable range, number of features present, and signals from control spots.

Supervised analysis of normalized gene expression data was conducted using the SAM algorithm. This algorithm was used to identify genes whose expression levels were significantly altered by influenza infection. We set the delta threshold in the SAM analysis to allow an acceptable false discovery rate (FDR) of 10%. We found that the expression levels for a total of 300 genes (representing 3.4% of the genes considered present on the chip) differed significantly between mock and infected samples (Table S2). Using the DAVID Bioinformatics Resources database, we annotated this signature using the gene ontology (GO) terms. This

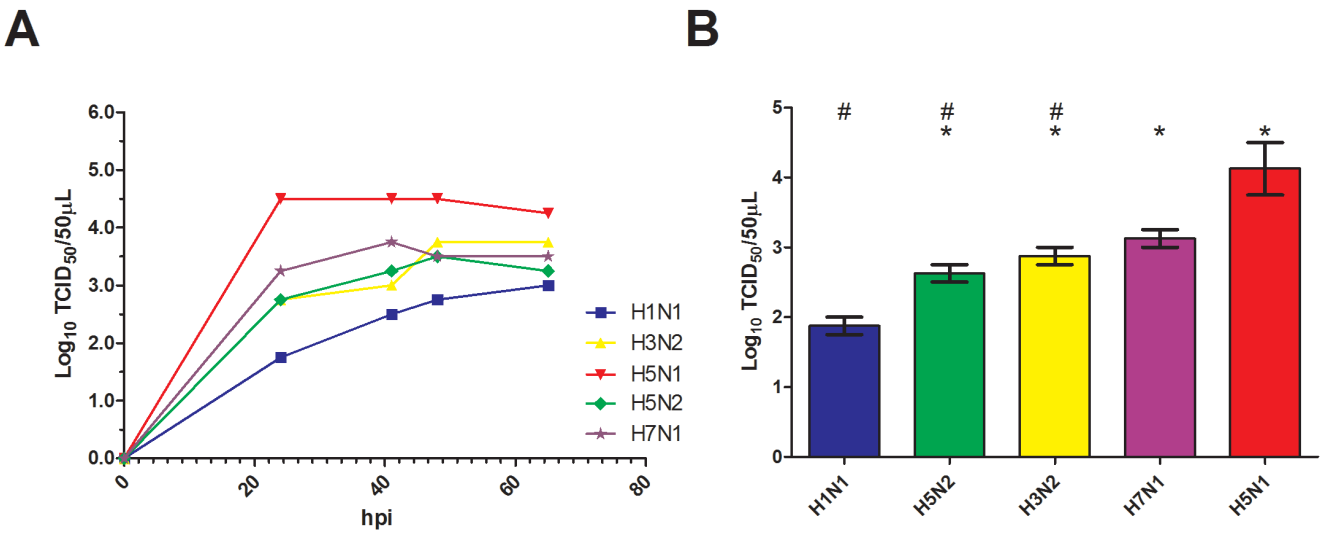


Figure 1. Comparison of viral replication kinetics between influenza viruses in A549 cells. A. A549 cells were infected at a moi of 0.1 with influenza virus A/New Caledonia/20/99 (H1N1), A/Moscow/10/99 (H3N2), A/Turkey/582/2006 (H5N1), A/Finch/England/2051/94 (H5N2), and A/Chicken/Italy/2076/99 (H7N1) and supernatants were collected at 24, 41, 48 and 65 hpi for end point titration assays. B. Each bar represents the mean of viral titers at 24 hpi for two independent experiments and titers were statistically analyzed by the one tailed Welch t-test; *: titers greater than H1N1 titers (p-value <0.05), #: titers lower than H5N1 titers (p-value <0.05).
doi:10.1371/journal.pone.0013169.g001

revealed an enrichment of genes related to various cellular processes such as protein complex biogenesis, membrane and microtubule organization, DNA metabolic and catabolic processes, cell proliferation regulation, cell cycle and cell death (Figure 2).

A subset of six genes with absolute fold changes in $\log_2$ (FCs) above 2 was selected to validate the microarray analysis by quantitative RT-PCR (RT-qPCR) analysis: *DNMT1*, *NTE* and *CAPN1* that were found downregulated in infected cells and *G1P2*,

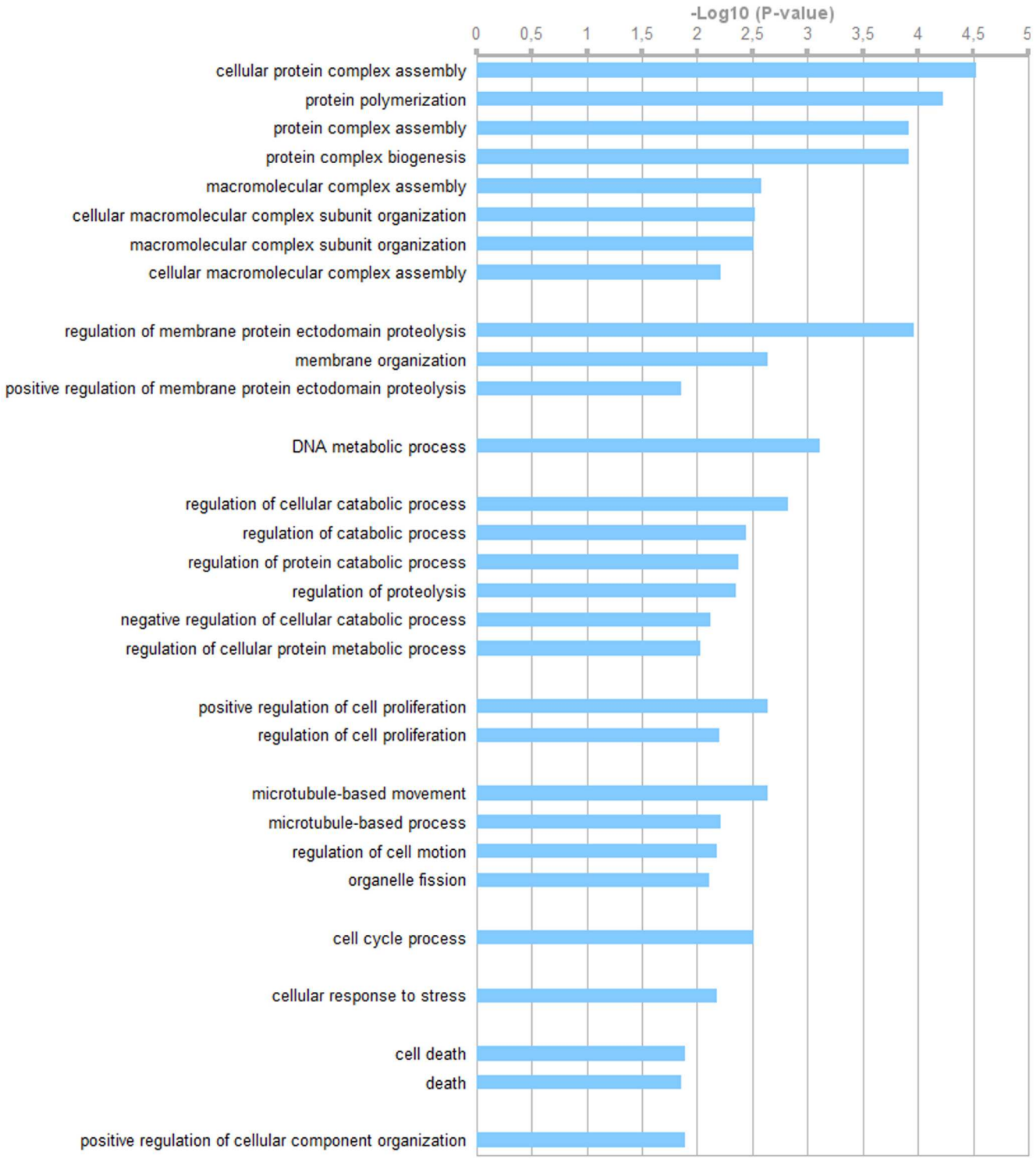


Figure 2. Gene enrichment analysis of the 300 genes of the infection signature. Discriminatory genes were analyzed by DAVID for associations with particular Gene Ontology terms. The negative Log_{10}(P-values) of enriched terms (plotted in bar) refer to how significant an association a particular ontology term has with the gene list. The 30 most significant biological process are grouped according to their biological meaning.
doi:10.1371/journal.pone.0013169.g002

OAS1 and *ICAM1* that were upregulated. The 6 genes were chosen at random among the most 20 dysregulated genes upon infection. This quantification was performed on new samples equivalent to those used for the microarray analysis. Figure 3 shows the confirmation by RT-qPCR of the microarray data. For each gene and each strain, microarray FCs are presented as a black boxplot and RT-qPCR results are depicted as a gray histogram. Results from RT-qPCR were in good agreement with the cDNA microarray analyses for five out of six genes tested. Indeed, except for *CAPN1* (p-value = 0.1), significant difference between infected and non infected cells was also observed in quantitative RT-PCR analysis (p-value <0.05, Welch t-test), similar to DNA microarray analysis. This result was acceptable considering that samples analyzed by RT-qPCR were different from those used in the microarray analysis.

To visually compare the changes in mRNA abundance for the 300 genes found to be influenced by influenza infection, hierarchical clustering analysis in both dimensions was performed. Results are depicted in the heatmap representation of Figure 4. Dendrograms indicate the correlation between samples and genes. We verified that mock samples were sorted together vs infected ones. The H1N1 samples co-clustered with the mock samples suggesting that infection with this strain induced few gene expression changes. We verified this result by conducting a virus-specific SAM analysis on the mock vs one virus samples. For a FDR of 10%, only 36 genes were found to be regulated by H1N1 infection in comparison to 2298 genes by H3N2, 1510 by H5N2, 3020 by H7N1 and 1455 by H5N1. The main difference between H1N1 and other viruses lay in the number of down-regulated genes during infection. Whereas H3N2, H5N1, H5N2 and H7N1 influenza viruses induced a down-regulation of most of the genes tested, a similar number of genes were down- and up-regulated by H1N1 (highlighted by the blue vertical line in the heatmap Figure 4). As H1N1 viral titer was lower at 24 hpi than titers of other viruses (Figure 1B), the scope of gene-expression changes induced upon infection correlated, at least partially, to the viral replication efficiency of the virus-cell system used in this study.

Interestingly, out of the 300 genes of the global infection signature, only 16 were upregulated in all infected cells. These 16 genes were associated to three GO biological process, including two related terms, "viral reproductive process" and "viral reproduction", that annotate genes encoding proteins involved in the virus life cycle. Two genes were associated to these terms: *ICAM1*, which is the major receptor for human rhinovirus [35], and *IRF7*, which activates the expression of Epstein-Barr Virus Latent Membrane Protein 1 [36] (p-value = 0.07 and 0.08, respectively). While *IRF7* has not been directly involved in influenza virus life cycle yet, *ICAM1* was recently identified as a proviral factor that may be co-opted by influenza virus [37]. The third associated biological process was the term "immune response" (p-value = 0.04) annotating 4 genes (*ICAM1*, *OASL*, *OAS1* and *CFD)*. Therefore, the upregulated genes were mostly associated with the immunological response. Besides, seven of the 16 genes were interferon stimulated genes (ISGs): *IFITM1*,

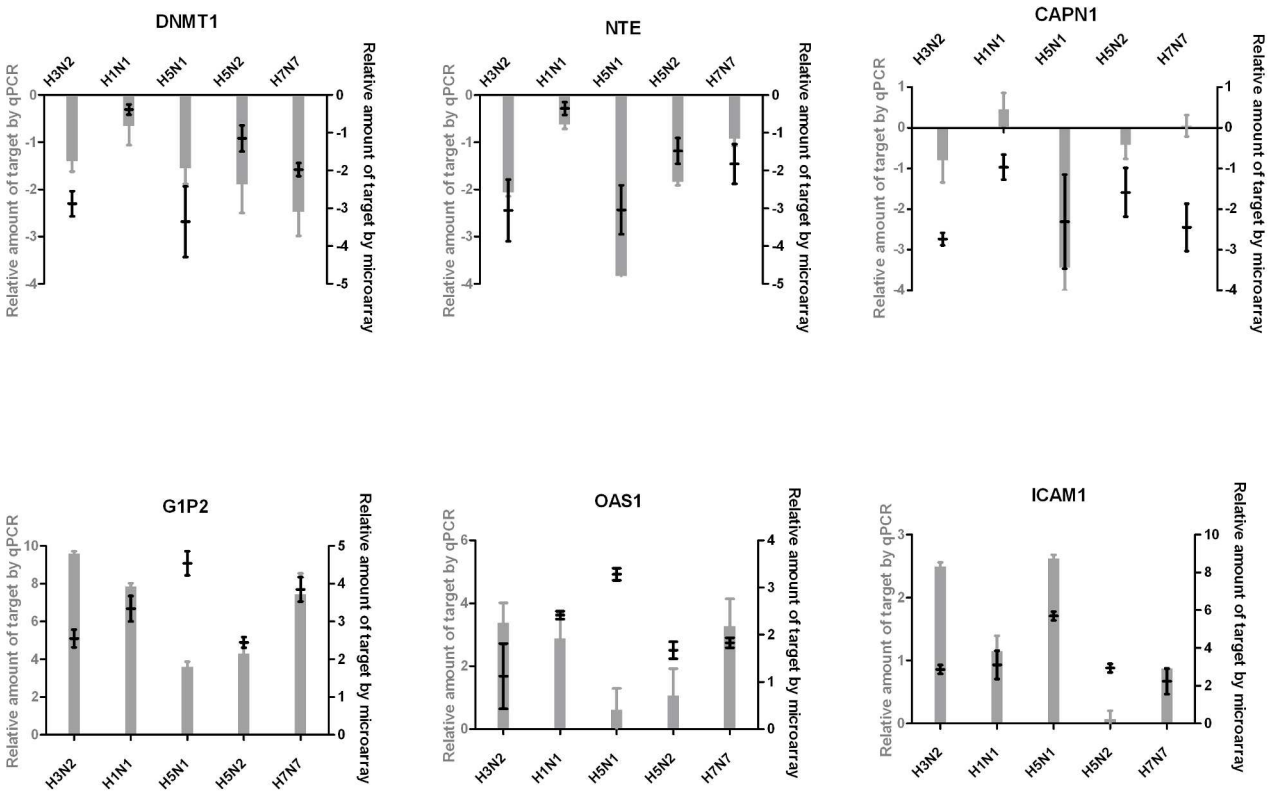


Figure 3. Validation of microarray results by real-time quantitative RT-PCR analysis. Expression of 6 genes from DNA microarray analysis (black cross) is compared with real-time quantitative PCR data (grey bar). Cross represent fold changes (mean ± S.D.) given by the microarray analysis = Log_2 (normalized amount of target gene in infected samples) - Log_2 (normalized amount of target gene in mock samples) (right). The formula used to determine the amount of target gene in infected cells by RT-qPCR, normalized to *GAPDH* and relative to mock is: $2^{-\Delta\Delta Ct}$ where Ct is the threshold cycle and $\Delta\Delta Ct$ = (Ct target infected – Ct gapdh infected) – (Ct target mock – Ct gapdh mock). Bar represented Log2 expression level (mean±S.D.) of target genes in two independent quantitative RT-PCR analysis (left).
doi:10.1371/journal.pone.0013169.g003

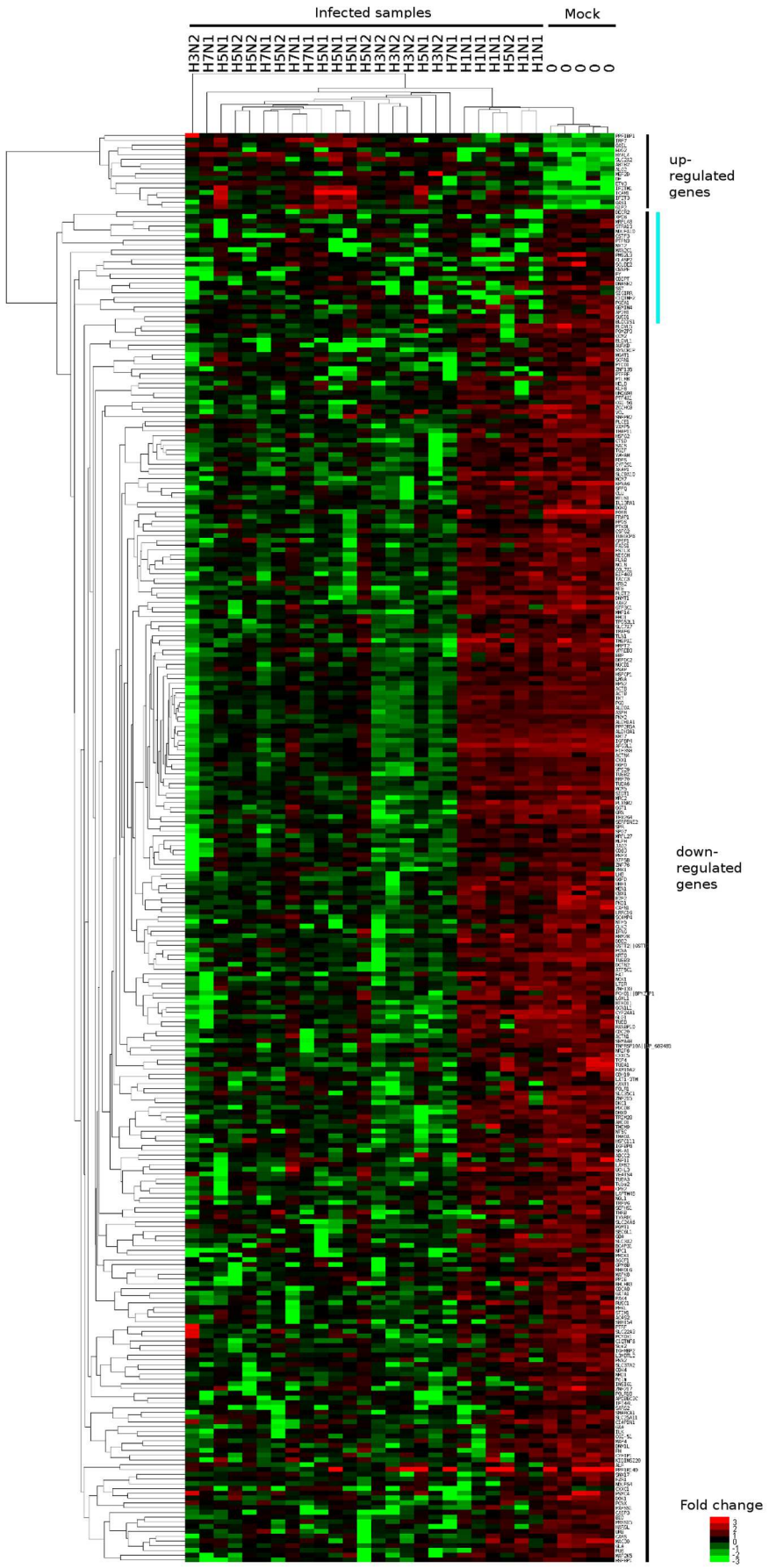

Infected samples
Mock
H3N2
H7N1
H5N1
H5N2
H5N2
H7N1
H5N2
H7N1
H7N1
H5N1
H5N1
H5N1
H5N2
H3N2
H3N2
H5N1
H3N2
H7N1
H1N1
H1N1
H5N2
H1N1
H1N1
0
0
0
0
0
up-regulated genes
down-regulated genes
Fold change
3
2
1
0
-1
-2
-3

Figure 4. Hierarchical clustering and heatmap of the 300 genes that discriminate mock and infected samples. Heatmap representing the expression levels of the 300 genes differentially expressed between infected and mock cells. Red squares indicate the expression levels above the median of the gene abundance across the sample and green squares the expression levels below. The median values were clustered hierarchically in both dimensions. Dendograms indicate the correlation between groups of samples (horizontal) and genes (vertical). Vertical lines in the right portion indicate the 2 distinct gene expression patterns: up and down-regulated genes during infection. The blue line stands for the genes down-regulated in samples infected with H1N1 influenza virus. Gene expression data for the 300 genes are reproduced in table S2.
doi:10.1371/journal.pone.0013169.g004

ICAM1, *IFIT3*, *OAS1*, *G1P2* (or *ISG15*), *IRF7* and *OASL*. These results were in accordance with previous studies showing the upregulation of immune response associated genes in samples infected in vitro and in vivo with various influenza viruses [9,11,12,37,38,39,40]. Gene expression levels in each group of samples are depicted in Figure S1. All ISGs were markedly more up-regulated in H5N1 infected cells than in other samples. This hyperstimulation has been described in other transcriptional studies [11,40,41] reinforcing the validity of the experimental cell-virus system developed in the present study.

2 *In silico* drug screening of the Connectivity Map

The Connectivity Map is a collection of genome-wide transcriptional expression data from cultured human cells treated with bioactive small molecules [13]. The associated website provides tools to find molecules connected to the query signature i.e. any list of genes associated with a biological test. The similarity of the query signature to each of the reference expression profiles is assessed and quantified by a normalized score, from -1 for a molecule that reverses the signature to +1 for a molecule which induces gene expression changes similar to the query signature.

Our strategy was to query the Connectivity Map with a list of genes differentially expressed in infected cells to find molecules that induced the opposite gene expression changes. We hypothesized that such molecules may influence host cell metabolism in such a way that effective viral replication would be altered.

A critical step in this screening was to define the query signature. As the number of upregulated genes was very low (5.3%) in the list of 300 genes defined by the analysis, a lack of specificity resulting from a loss of information for up-regulated genes could be introduced in drug selection if the signature was not corrected for this bias. By selecting genes with the most drastic changes in level of expression (fold change in $\log_2 > 2$ or <-2), we were able to define a signature of 20 genes for influenza A virus infection with similar amounts of those up and down regulated (Figure 5A, Table 1).

By querying the connectivity map with this concise signature, we obtained c-scores for 6100 instances, representing more than 1000 molecules in various conditions [14]. We selected those associated with the most strongly anticorrelated signatures (negative enrichment) and which had a p-value less than 0.5% (Figure 5B and 5C). Applying this filtering step left us with eight candidate molecules: harmol, rilmenidine, brinzolamide, ribavirin, calcium folinate, 2-aminobenzenesulfonamide, merbromin and midodrine (from the most negatively correlated to the least negatively correlated drug). The relevance of our selection was supported by the fact that ribavirin, an already known influenza virus inhibitor, was identified with a negative enrichment of -0.83 and a pvalue of 0.00157. Except for the topical antiseptic merbromin, the other selected molecules have various therapeutic indications (depicted in Table S3) but are not referenced as antivirals.

Graphs in Figure 5C report how the different genes of the infection signature behave in the expression profile of the selected molecules. Although the genes down-regulated during infection are generally up-regulated in response to the molecule and conversely the up-regulated genes of the signature are globally down-regulated by the molecule, none of the molecules available in this data bank were able to completely reverse the infection signature.

3 Evaluation of the antiviral potency of the selected drugs on H3N2 viral growth

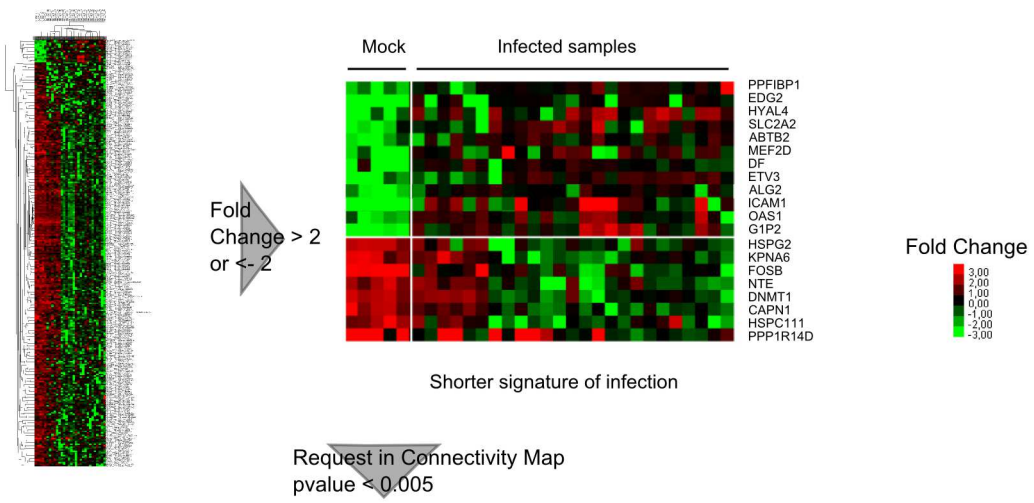
We assessed the effect of the eight selected molecules on influenza replication in vitro. Cell viability, as assessed by the neutral red assay, and viral growth, as quantified by a neuraminidase (NA) activity test, were conducted in parallel. Before using the NA activity test as an indirect measurement for viral impairment, we checked firstly that the different influenza viruses used in this study had sufficient neuraminidase activities to be quantified using this method (Figure S2A). For all tested viruses and for a signal to background ratio between 2 and 70, the fluorescence was proportional to the amount of virus present ($TCID_{50}$/mL). During the evaluation of the drug panel, all signal to background ratios were included between 2 and 70. Secondly, we controlled that the different molecules did not inhibit the enzymatic activity of NA to be sure that a drop in RFU would only reflect a drop of viral titer. While concentrations of merbromin above 50 μM and harmol above 500 μM inhibited NA activity, incubation of the virus with increasing concentrations of the molecules otherwise resulted in no inhibition (Figure S2B). Therefore, for these two molecules below these concentrations and for other molecules of the drug panel, viral growth can be assessed by a neuraminidase test.

Evaluation of the drug panel was first conducted on influenza A/Moscow/10/99 (H3N2) virus. A549 cells were incubated with increasing concentrations of the molecule for 6 h before infection. This time was chosen based on the duration of treatment indicated in the Connectivity Map to obtain similar cellular response before infection [14]. Infection was allowed to proceed for 65 h which represents multiple cycles of infection, however similar results were observed at 24 and 48 hpi (Figure S3).

The viability data of five independent experiments are given in Figure S4. The 50% cytotoxic concentrations (CC_{50}) were determined by regression analysis. The CC_{50} of calcium folinate, 2-aminobenzenesulfonamide and midodrine could not be determined since none of these molecules was cytotoxic at the highest tested dose.

The effect of each of the molecules on viral growth was tested using the H3N2 virus at a moi of 0.2 and 2. Dose-response curves were fitted by regression analysis (Figure S5) and used to determine the 50% effective concentration (EC_{50}) of each molecule if at least one response was inferior to 50%. Selective indexes (SI) were calculated as CC_{50}/EC_{50} and used to classify selected molecules as inactive ($SI<2$), weak inhibitors ($2<SI<10$), moderate inhibitors ($10<SI<50$) and strong inhibitors ($SI>50$) (Figure 5). In agreement with previous observations [42], we noted that SI were dependent on the moi, since molecules are more effective at lower moi. In our conditions, at a moi of 0.2, two molecules (calcium folinate and 2-aminobenzenesulfonamide) were ineffective, two (harmol and merbromin) were weak inhibitors, two (brinzolamide and midodrine) were moderate inhibitors and one (ribavirin) was a strong inhibitor. At a moi of 2, whereas brinzolamide was reclassified as a weak inhibitor, the other molecules remained in the same class despite their SI being

A



B

by name | by name and cell line | by ATC code

rank	cmap name	mean	n	enrichment	p	specificity	% non-null
1	sirolimus	0.402	44	0.377	0.00000	0.1807	72
2	trichostatin A	0.334	182	0.316	0.00000	0.6967	67
3	tanespimycin	0.363	62	0.299	0.00000	0.3731	66
4	thioridazine	0.456	20	0.528	0.00004	0.2055	80
5	sulfametoxydiazine	0.637	4	0.895	0.00010	0.0000	100
6	DL-thiorphan	0.757	2	0.989	0.00016	0.0000	100
7	harmol	-0.667	4	-0.889	0.00034	0.0109	100
8	methylbenzethonium chloride	0.549	6	0.767	0.00042	0.0157	83
9	alvespimycin	0.477	12	0.548	0.00052	0.0460	91
10	pyrvinium	0.548	6	0.747	0.00070	0.0791	100
11	suloctidil	0.623	4	0.842	0.00101	0.0454	100
12	rilmenidine	-0.582	4	-0.844	0.00105	0.0000	100
13	brinzolamide	-0.461	4	-0.836	0.00125	0.0000	100
14	ribavirin	-0.549	4	-0.830	0.00157	0.0132	100
15	calcium folinate	-0.363	5	-0.740	0.00236	0.0229	60
16	LY-294002	0.281	61	0.226	0.00306	0.5705	59
17	latamoxef	0.603	3	0.865	0.00453	0.0145	100
18	2-aminobenzenesulfonamide	-0.304	4	-0.782	0.00462	0.0139	50
19	merbromin	-0.453	5	-0.706	0.00477	0.0325	80
20	midodrine	-0.326	5	-0.705	0.00491	0.0752	60

Drugs negatively connected to infection

C

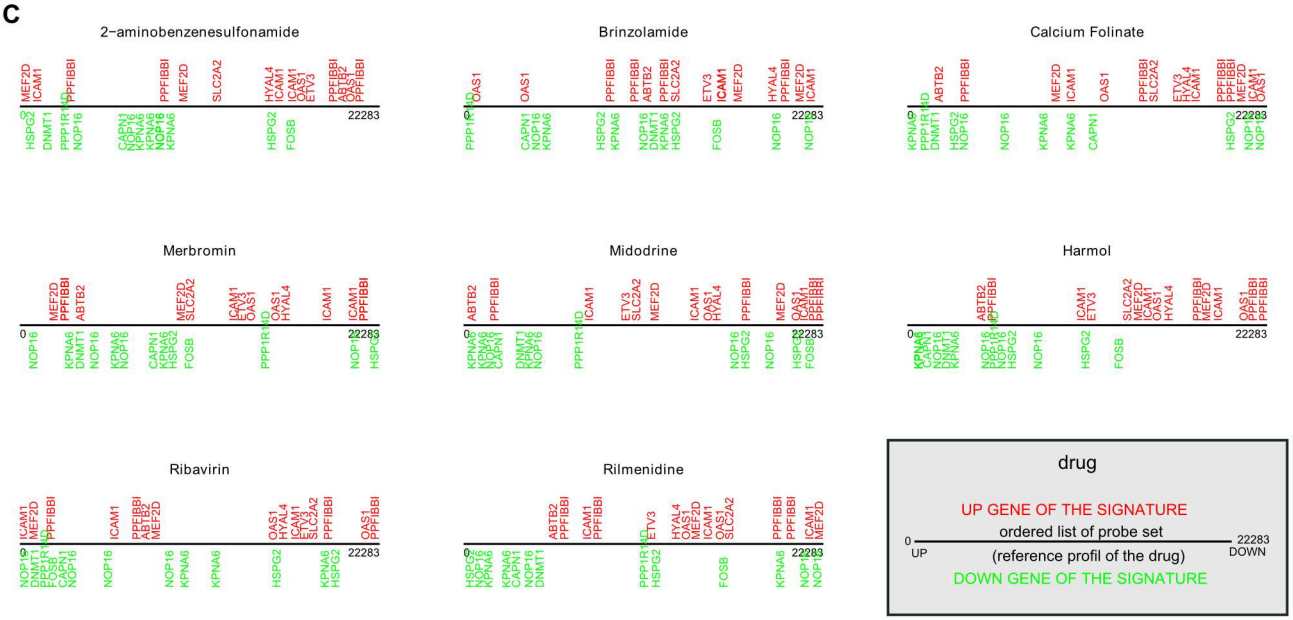


Figure 5. Gene expression-based screening identifies eight potential antiviral molecules. A. List of cellular genes chosen to query the Connectivity Map. A circumscribed signature of infection was derived from the 300 genes discriminating mock and infected samples by selecting genes with a fold change > 2 or < -2. The fold change is defined as the ratio of the mean gene expression in infected samples to the mean for the corresponding mock infection in $\log_2$. This selection resulted in a list of 20 genes, 12 being up-regulated during infection and 8 down-regulated (see Table 1). These genes constituted the signature used to query the online database Connectivity Map. B. Drugs with significant enrichment to influenza virus infections in the Connectivity Map. Significance cut-off was set at p-value <0.005. The permutation p-value estimates the likelihood that the results would be observed by random chance. Mean: the arithmetic mean of the connectivity scores for the post-dose changes (or instances) by the given molecule; n: the number of instances of a given molecule in the CMAP database; Enrichment: a measure of the enrichment of those instances in the order list of all instances, Positive enrichment scores are of interest if perturbagens inducing the biological state represented by the signature used to produce the result are sought. Likewise, if reversal or repression of the biological state encoded in the query signature is required, perturbagens with negative enrichment scores are of interest. The specificity value is defined as the frequency at which the enrichment of a set of instances equals or exceeds that of the same set of instances in queries executed on 312 published, experimentally derived signatures using the Molecular Signatures Database. Lower values are associated with a greater specificity; the non null percentage represents a measure of the support for the connection between a set of instances and a signature of interest based on the behavior of the individual instances in that set. C. 8 molecules are negatively connected to influenza virus infection (p-value <0.005). A graphical representation of the location of the signature of infection is depicted for each molecule, taking the instance with the most negative connectivity score of each molecule. The x-axis represents the genes of the expression profile of the molecule, rank ordered according to their differential expression relative to the control. The location of each gene of the infection signature is appreciated along the x-axis.
doi:10.1371/journal.pone.0013169.g005

weaker. As an example, the CC_{50} for midodrine was superior to 4250 μM and EC_{50} was comprised between 322 μM (moi of 0.2) and 532 μM (moi of 2). Concerning rilmenidine, which was dissolved in DMSO, it was not possible to conclude on an effect. DMSO has previously been shown to be cytotoxic and to inhibit influenza infection above 4% (vol/vol) [43] however it is still used as a major solvent for molecules in high-throughput screening. In this study, the CC_{50} for DMSO was 2.9% (v/v) - the concentration used to obtain 1550.7 μM of rilmenidine- and the EC_{50} was comprised between 1.0% (moi of 0.2) and 1.8% (moi of 2). The EC_{50} of rilmenidine was significantly different from that of DMSO at a moi of 2 ($p = 2e-4$) but not at a moi of 0.2 ($p = 0.85$).

However, even if this molecule is considered ineffective against the H3N2 influenza virus, we did obtain a very high confirmation rate (62.5%) in comparison with the hit rate of classical high-throughput screening (3.2% in [43]). This clearly indicates that

Table 1. Genes of the circumscribed signature of infection.

Gene	Gene symbol	I.M.A.G.E. CloneID	Fold Change
Transcription Regulation			
Ets variant gene 3	**ETV3**	1473929	7.24
Myocyte enhancer factor 2D	**MEF2D**	4209	5.13
FBJ murine osteosarcoma viral oncogene homolog B	**FOSB**	79022	−2.69
DNA (cytosine-5-)-methyltransferase 1	**DNMT1**	768241	−2.06
Ankyrin repeat and BTB (POZ) domain containing 2	**ABTB2**	204456	2.28
Immune response			
ISG15 ubiquitin-like modifier, G1P2	**ISG15**	742132	3.48
2′,5′-oligoadenylate synthetase 1, 40/46kDa	**OAS1**	666703	2.07
Complement factor D (adipsin), DF	**CFD**	666128	2.99
Cell surface protein			
Patatin-like phospholipase domain containing 6, NTE	**PNPLA6**	431990	−2.01
Heparan sulfate proteoglycan 2	**HSPG2**	3339	−2.23
Solute carrier family 2 (facilitated glucose transporter)	**SLC2A2**	207963	2.23
PTPRF interacting protein, binding protein 1 (liprin beta 1)	**PPFIBP1**	263094	2.04
Intercellular adhesion molecule 1 (CD54)	**ICAM1**	3383	3.36
Lysophosphatidic acid receptor 1	**LPAR1**	505524	3.63
Enzyme			
Hyaluronoglucosaminidase 4	**HYAL4**	668088	2.90
Asparagine-linked glycosylation 2 homolog	**ALG2**	150443	2.35
Calpain 1, (mu/I) large subunit	**CAPN1**	70555	−2.20
Protein phosphatase 1, PP1	**PPP1R14D**	140525	−3.13
Nuclear protein			
Karyopherin (importin) alpha 6	**KPNA6**	16650	−2.21
Hypothetical protein HSPC111	**NOP16**	505242	−2.08

Fold Change = $\log_2$(Inf/Mock).
doi:10.1371/journal.pone.0013169.t001

our in silico screening was effective and strongly supports its power at selecting the antivirals: harmol, merbromin, brinzolamide, midodrine and ribavirin.

4 Antiviral effects of most of the molecules are due to an action on cells rather than on viruses

Molecules selected by the in silico screening were chosen from the Connectivity Map based on the gene expression changes they induce in treated cells. To check that the antiviral properties of the five efficient molecules were actually mediated by an action on cells and not by an indirect effect on the virus, we conducted two assays in parallel in which either the cells or the H3N2 virus were preincubated with a series of concentration of the molecules. The efficiencies of infection were estimated by measuring the neuraminidase activity associated to cells at an early time of infection. In the preincubated cells assay, cells were in contact with molecules for 14 hours before being infected with H3N2 virus without any drugs. As the cells were washed twice before infection, we assumed that the virus should not be in contact with the molecules during infection. Thus the molecules should not alter the viral structure nor change parameters playing a direct role on viral entry (as the extracellular pH for example). Consequently an inhibition of infection in this assay would mean that the molecule had an effect on cells. In contrast, in the preincubated virus test, the viral stock was treated with the molecules during 14 hours while the cells were in contact with molecules though after dilution and for only 15 minutes during infection. We assumed that this exposure time and molecule concentrations were too low to induce any effect on the cells. If a molecule should inhibit viral growth by altering the functional properties of the virus (viral structure or surface glycoprotein), infection would be inhibited in the preincubated virus condition but not in the preincubated cells one.

Results of both tests for the five efficient molecules are depicted in Figure 6. After preincubating the viral stock with the molecules, a few infection efficiencies were significantly different of the control (p-value <0.05, two tailed Welch t-test). However, except for merbromin, infection efficiencies after virus preincubation were included between 64% (for ribavirin $c = 400\ \mu M$) and 110% of the control (for rilmenidine $c = 8\ \mu M$). Therefore, the different drugs exerted very limited effects on the virus. In contrast, statistically significant inhibitions of infection efficiency were noted after cells preincubation with each molecule at higher concentrations (above 10 μM for brinzolamide, 40 μM for harmol, 1 μM for merbromin, 140 μM for midodrine, 160 μM for ribavirin and 80 μM for rilmenidine). Infection efficiency decreased to 23% for brinzolamide (100 μM), 5% for harmol (800 μM), 2% for merbromin (250 μM), 40% for midodrine (1400 and 4250 μM), 26% for ribavirin (800 μM) and 23%3 for rilmenidine (1600 μM). We concluded from these tests that the antiviral effect of these molecules is due to an action on cells rather than on the virus. Merbromin on the other hand inhibited viral infection in both assays. This was not surprising since this molecule is a topical antiseptic known to inactivate influenza viruses [44]. However, our results indicate that this molecule may also inhibit viral replication through a cellular effect.

5 None of the molecules which are positively correlated to the infection signature, impaired H3N2 influenza viral growth

In order to control that the antiviral effect of the molecules is specifically associated with inversion of the infection signature, we assessed the effect of some molecules positively correlated to the signature. Seven drugs, alvespimycin, DL-Thiorphan, latamoxef, methylbenzethonium chloride, pyrvinium, sulfameter (or sulfamethoxydiazine) and sulodictil, were chosen according to the following criterion: p-value $<0.5\%$, mean > 0.35 and a specificity <0.1 (Figure 5B).

Viability and viral growth assays were performed on A549 cells infected with H3N2 virus at a moi of 0.2 and 2, as described for negatively correlated drugs. Dose-response curves (Figure S6 and S7) were used to determine CC_{50} and inhibitory EC_{50} (Table S4). In these conditions, inhibitory SI were lower than 2, or than SI of DMSO for DL-Thiorphan and Sulodictil. Thus none of the positively correlated drugs inhibited viral replication at both moi. In contrast, four drugs (alvespimycin, methylbenzethoniumchloride, pyrvinium and sulodictil) enhanced viral production at a moi of 0.2. Increase of viral titers ($TCID_{50}$/mL) was up to 2 $\log_{10}$ and was statistically significant for alvespimycin, methylbenzethoniumchloride, and sulodictil 40 μM (p-value <0.05, two tailed Welch t-test). Therefore, these results strengthen our hypothesis that modulation of host cell transcription may have an impact on viral replication.

6 Some antivirals are effective against a broad range of influenza A virus strains, including the pandemic H1N1 influenza virus

We hypothesized that one advantage of our gene-expression based screening is that the selected molecules would have an activity against various influenza A viruses. Indeed, since we selected a gene signature of infection common to different human and avian strains, we assumed this as a prevailing cellular response to many influenza viruses. Therefore, we tested the effect of the selected molecules on the viral growth of the different strains used for the initial microarray analysis, i.e A/New Caledonia/20/99 (H1N1), A/Turkey/582/2006 (H5N1), A/Finch/England/2051/94 (H5N2), and A/Chicken/Italy/2076/99 (H7N1). Two independent assays in duplicate (4 replicates in total) for each virus were conducted in the conditions previously defined for the H3N2 virus. EC_{50} and SI were determined for each molecule and are summarized in Table 2, Table 3 and Figure 7. Molecules that inefficiently inhibited growth of the H3N2 strain (2-aminobenzenesulfonamide and calcium folinate) were also inefficient against other tested viruses. Conversely, the strongest H3N2 inhibitor, ribavirin, was also classified as a strong inhibitor of all viruses tested. However, ribavirin obtained different SI depending on the viral strain tested, allowing the viruses to be classified according to their sensitivities to this molecule: H3N2 > H5N2 and H1N1 > H7N1 > H5N1. Other drug screening tests carried out previously on MDCK cells (with an moi of 0.001) had already reported a higher sensitivity of H3N2 viral strains compared to H1N1 [42,45]. In our tests, ribavirin EC_{50} was comprised between 6 μM (for H1N1 and H3N2 with an moi of 0.2) and 632 μM (for H5N1 with an moi of 2) in concordance with previously published results [42,45]. Midodrine, the second most active molecule against the H3N2 strain, also showed an antiviral effect against both H1N1 (SI >2.7 for moi 2 and SI >142.5 for moi 0.2) and H5N2 (SI >2.5 for moi 2 and SI >8.9 for moi 0.2) viral strains. Brinzolamide was a moderate inhibitor of human H3N2 and H1N1 influenza viruses and a weak inhibitor of avian H5N2 and H7N1 influenza viruses. Harmol weakly inhibited all viruses tested, as did merbromin the EC_{50} for which were near to 50 μM, a concentration noted to interfere with the neuraminidase activity test. Finally, rilmenidine had an obvious antiviral effect on the H1N1 strain. Some of the molecules identified by our approach were therefore able to inhibit viral growth of all the viruses used to define the gene expression signature of infection.

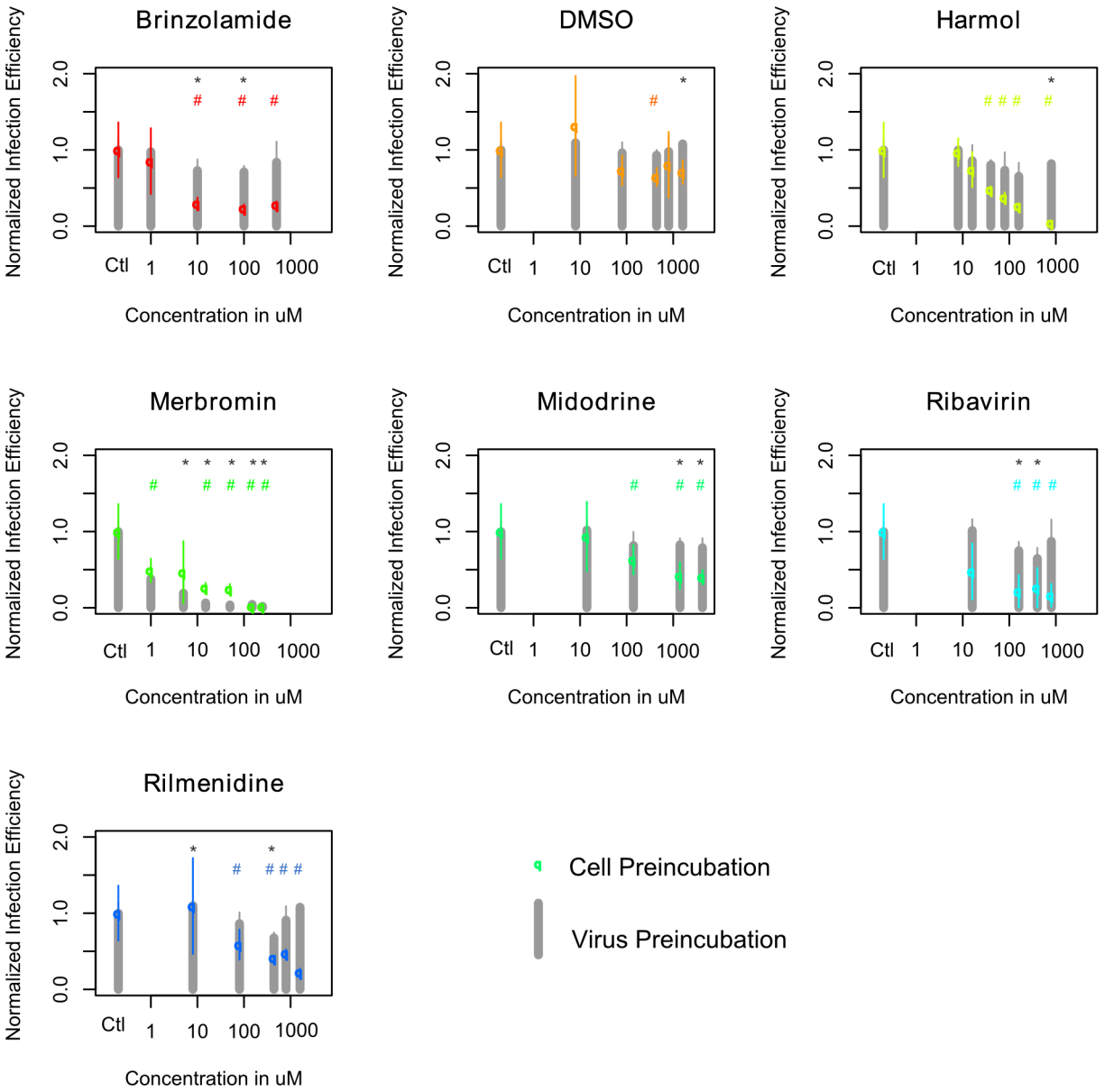


Figure 6. Antiviral effects of most of the molecules are due to an action on cells rather than on the virus. In the 'Cell Preincubation' assay, A549 cells were treated with increasing concentrations of the molecules or solvent for 14 h and washed twice before incubation with H3N2 influenza A virus at an moi of 7 during 15 min and infection for 5 h in medium without drug. In the 'Virus Preincubation' assay, H3N2 viral stock was treated with increasing concentrations of molecules or solvent for 14 h and diluted 12 times before incubation with cells during 5 min and infection for 5 h. In both assays, the number of infected cells was estimated with a neuraminidase test after cell lysis with Triton 1X. The normalized infection efficiency was calculated as RFU in treated cells/RFU in control cells. Values represent the mean of two independent experiments (+/− standard deviation). Results of the preincubated cell test are plotted with colored dots and results of the virus preincubated assay are depicted with grey bars. * and # indicate statistically significant differences of infection efficiency in comparison to untreated control cells (Welch two-sample t test, pvalue <0.05), for the preincubated cell test and for the preincubated virus test, respectively.
doi:10.1371/journal.pone.0013169.g006

To determine if this strategy identified broadly effective influenza antivirals which could be active against emerging influenza viruses, we tested their effect on the viral growth of the recent pandemic H1N1 virus (referred to as H1N1 SOIV) (Figure S8). Interestingly, in comparison with A/New Caledonia/20/99 (H1N1) virus, a weak to moderate antiviral effect was observed for 2-aminobenzenesulfonamide whereas rilmenidine was ineffective. The other molecules had comparable effects on the two H1N1 virus strains, with brinzolamide, midodrine and ribavirin being the most effective antivirals. The EC_{50} of ribavirin were comprised between 61 µM (for an moi of 0.2) and 292 µM (moi of 2) revealing a resistance to this molecule that was 4 (moi 2) to 10 (moi 0.2) times more in the H1N1 SOIV strain compared to the H1N1 strain (pvalue = 3.6e−6 and 0.0012 respectively).

We compared drug sensitivities to viral growth curves of different viruses after infection of A549 cells at two moi (Figure S9). Viruses with good replication efficiencies and the faster kinetics (H5N1 and H7N1) were the most resistant to the drug panel. In contrast, selected antivirals had a better effect on delayed replication viruses (H1N1 and H1N1 SOIV). Drug sensitivities

Table 2. Potency of the inhibitors against human influenza A viruses in A549 cells.

		H1N1 moi 2	H1N1 moi 0.2	SOIV moi 2	SOIV moi 0.2	H3N2 moi 2	H3N2 moi 0.2
	CC50 (μM)	EC50 (μM)	EC50 (μM)	EC50 (μM)	EC50 (μM)	EC50 (μM)	EC50 (μM)
2-Aminobenzenesulfonamide	>6900	>690	>690	1821.67	533.83	>6900	6020
Brinzolamide	665.51	435.06	28.95	166.65	21.55	131.14	23.00
Calcium Folinate	>2400	>800	>800	>2400	>2400	>2400	>2400
Harmol	94.98	60.36	20.53	53.74	6.76	20.07	11.68
Merbromin	103.80	29.55	61.98	37.53	>50	32.38	42.87
Midodrine	>4250	1566.7	29.82	929.96	108.79	321.99	531.53
Ribavirin	29528	73.79	6.63	291.52	61.08	12.53	5.79
Rilmenidine	1125.3	506.22	24.22	1073	864.48	475.55	489.87
DMSO	1550.7	>1600	173.63	606.40	344.03	957.2	554.39
p-value (Rilmenidine, DMSO)	p = 0.0026	p = ND	p = 7e−04	p = 0.6128	p = 0.3552	p = 2e−4	p = 0.88

CC_{50}: molecule concentration of 50% cytotoxicity; EC_{50}: molecule concentration of 50% inhibition of viral replication; p-value: t-statistic for testing Rilmenidine values equal to DMSO values.

doi:10.1371/journal.pone.0013169.t002

therefore partially correlated with viral growth kinetics. However, some strain specificity may also account for drug sensitivities. Indeed, H3N2 virus was one of the most drug sensitive virus, while replicating as efficiently than H7N1 virus (but a bit slower).

To conclude, five molecules out of the eight potential molecules selected by our in silico screening inhibited viral growth of the H1N1 SOIV, a virus that was unknown when we first defined the signature of infection and queried the Connectivity Map. These results are promising and strongly indicate that this approach identifies molecules with a broad anti-influenza spectrum of activity.

Discussion

The virally induced gene-expression signature

Influenza infection induces various intracellular signaling cascades and important downstream gene expression host-cell modifications [46]. Despite their host-range restriction that may reflect the better adaptation to host factors [47], all influenza A viruses can infect the same cells in vitro, prompting us to assume that they may hijack common cellular proteins for their own replication. This is the first study to compare the cellular gene expression modifications induced by five different influenza A virus subtypes. As already described in previous transcriptional in vitro [40] and in vivo studies [11,12], we found that H5N1 infection induced a strong upregulation of interferon response genes. This sustained hyperinduction has been correlated with the high virulence of this virus in animal models [11,12]. In patients, H5N1 infection results in a massive production of cytokines and chemokines, referred to as the cytokine storm, which could be responsible for the severity of the disease [48]. Here we observed that H5N1 induced the expression of more, and to a greater extent, inflammatory/immune response genes than any of the other subtypes. Molecular mechanisms supporting the higher activation of interferon signaling by H5N1 in comparison with other subtypes remain undetermined. In contrast, we found that A/New Caledonia/20/99 (H1N1) infection leads to the smallest change in gene expression at 24 hpi. One could speculate that H1N1 virus, as a human influenza virus, would be well adapted to

Table 3. Potency of the inhibitors against avian influenza A viruses in A549 cells.

		H5N1 moi 2	H5N1 moi 0.2	H5N2 moi 2	H5N2 moi 0.2	H7N1 moi 2	H7N1 moi 0.2
	CC50 (μM)	EC50 (μM)	EC50 (μM)	EC50 (μM)	EC50 (μM)	EC50 (μM)	EC50 (μM)
2-Aminobenzenesulfonamide	>6900	>690	>690	>690	>690	>690	>690
Brinzolamide	665.51	>500	>500	232.40	145.61	551.05	139.19
Calcium Folinate	>2400	>800	>800	>800	>800	>800	>800
Harmol	94.98	77.49	43.36	4.06	13.45	49.35	27.96
Merbromin	103.80	44.99	45.98	36.45	21.93	43.70	46.74
Midodrine	>4250	>4250	3896.3	1670.7	475.58	>4250	>4250
Ribavirin	29528	632.27	48.36	16.40	6.71	32.39	12.47
Rilmenidine	1125.3	2037.7	1466.8	514.32	358.69	1050.4	777.85
DMSO	1550.7	>1600	1829.6	733.6	564.1	1109.3	801.08
p-value (Rilmenidine, DMSO)	p = 0.0026	p = ND	p = 0.40	p = 2e−04	p = 0.1881	p = 0.53	p = 0.9

CC_{50}: molecule concentration of 50% cytotoxicity; EC_{50}: molecule concentration of 50% inhibition of viral replication; p-value: t-statistic for testing Rilmenidine values equal to DMSO values.

doi:10.1371/journal.pone.0013169.t003

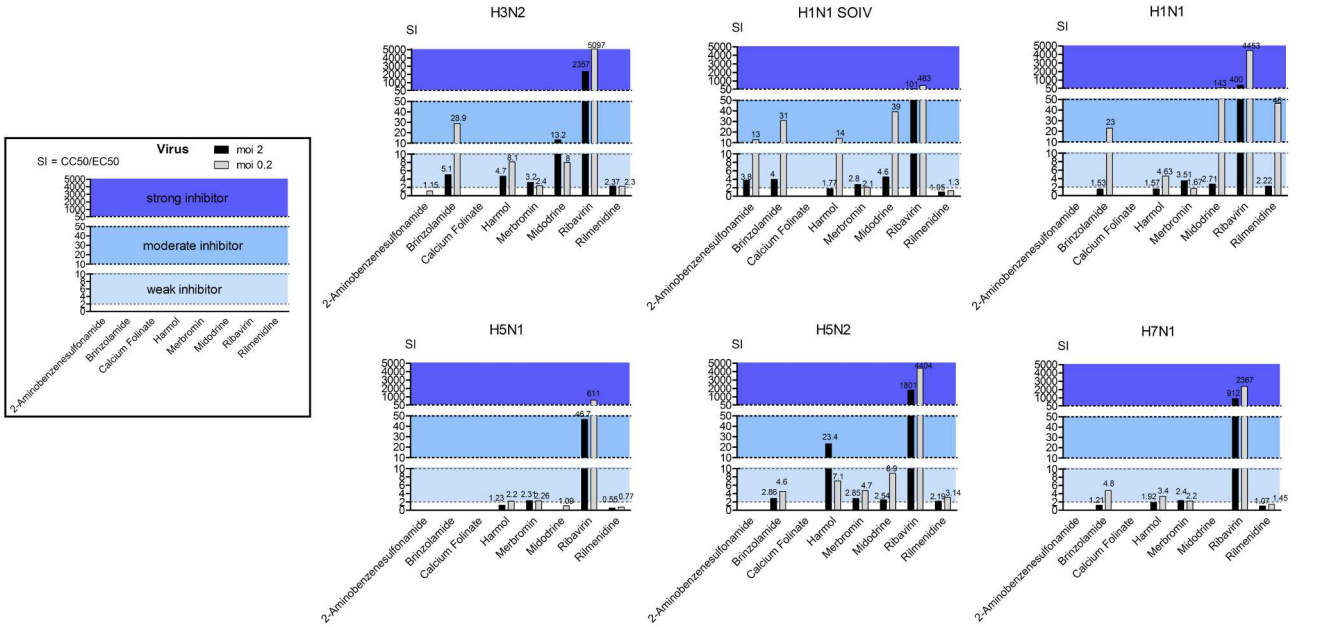


Figure 7. Antiviral potency of the 8 negatively correlated molecules on different influenza A viruses (moi 0.2 and moi 2). Selective indexes (SI) were calculated as CC_{50}/EC_{50} (Tables 2 and 3) and used to classify molecules as inactive (SI<2), weak inhibitors (2<SI<10), moderate inhibitors (10<SI<50) and strong inhibitors (SI>50).
doi:10.1371/journal.pone.0013169.g007

human A549 cells and could replicate in these cells with basal level of proteins, thus without having to induce much gene-expression changes. However a well adapted virus would efficiently replicate in these cells. We performed replication kinetics in A549 cells with the different viruses and observed that H1N1 virus grew to lower titers than other viruses. Two hypothesis can be formulated to explain the correlation between the weak growth of H1N1 virus and the few changes of host transcription. Either the reduced virus replication efficiency of H1N1 virus is responsible for the lower host response. This is supported by previous study where the replication efficiency of the virus-cell system accounts for the level of the host innate immune response [49]. Or it is also possible that H1N1 viral replication is impaired because of its inability to modulate the host response, especially to induce proviral pathways. This hypothesis is based upon previous demonstration that stronger virus-induced MAPK activation resulted in higher viral replication efficiency [50].

Nevertheless, beyond these subtype-specific profiles, we were able to identify a list of 300 genes differentially expressed in both mock and infected samples. Strikingly, only about 5% of these genes were upregulated. A similar imbalance has previously been observed in other transcriptional profiles of infected cell lines [40,51]. One could hypothesize that this may reflect the virally-induced cellular arrest of protein expression and could be due to the 5′cap snatching and subsequent degradation of cellular mRNA [52] and/or the inhibition of processing and export of cellular mRNA by NS1 [53]. Nevertheless these downregulated genes represented only 3.3% of the total number of genes detected, suggesting that a selective inhibition of their expression may occur during infection. The downregulated genes are implicated in different cellular processes such as ATP binding, regulation of translation, cellular protein complex assembly, glucose metabolic processes, cell cycle and apoptotic mitochondrial changes. On the other hand, the 16 genes found upregulated are specifically associated with innate cellular immunity. Seven of these are induced by interferon: *OAS1*, *ISG15*, *IRF7*, *OASL*, *ICAM1*, *IFITM1*, and *IFIT3*. These 7 ISGs have already been found upregulated together with other interferon genes upon H1N1 PR8 endothelial primary cell cultures infection [37] (Table S5). We also found an upregulation of *CFD*, a gene coding for a component of the alternative complement pathway. Complement is an important player in immunity and is induced by influenza infection [11,54,55]. Other induced genes of the infection signature determined in this study have never before been associated with influenza infection. They include *ETV3* which encodes a transcriptional repressor [56] that could be partially responsible for the downregulation of other genes belonging to the signature.

Signature use for drug screening

Here we identified a list of genes whose expression is significantly altered during infection with different human and avian influenza virus subtypes. Since the outcome of infection appeared successful in our experimental conditions, it can be concluded that such a virally-induced cellular environment is favorable for virus replication. We therefore hypothesized that any molecule able to inverse the infection signature should be harmful to influenza virus replication. In contrast to many published transcriptomic studies [1,11], we did not focus on a particular gene with a known function or large annotation that can be assumed to have a link with viral infection. To conduct the in silico screening, we filtered the infection signature genes according to their level of expression and selected the twenty most differentially expressed (with statistical significance) between mock and infected cells. We therefore took into account all of the information retrieved from the transcriptional analysis, which was a major advantage when using the Connectivity map. We selected eight molecules which induced gene expression modifications which anti-correlated with the infection signature. The hit-rate for this in silico screening was 0.53%.

Our experimental strategy presented several limitations: (i) we used a nylon microarray containing only 8000 genes thus meaning that the transcriptional profile of infected cells is incomplete; (ii)

this profile was assessed for an established cell line, A549, which is different from those used in the Connectivity Map (MCF7, HL60 and PC3); (iii) the Connectivity Map contains data for only 1000 molecules and none of the molecules we identified was able to induce a full inversion of the infection signature (Figure 5C). Despite these limitations, seven molecules out of the eight selected by the in silico screening presented an antiviral effect on at least one of the tested viruses (87.5% of molecules confirmed). 2-aminobenzenesulfonamide and rilmenidine had only a modest antiviral effect on one specific virus (respectively H1N1 SOIV and H1N1 New Caledonia). Harmol and merbromin were weak inhibitors of most of the tested viruses. Brinzolamide and midodrine were weak to moderate inhibitors of most of the tested viruses. As expected, ribavirin was a strong inhibitor of all tested viruses. In light of these results, we conclude that we have identified a common signature whose partial inversion is strong enough to inhibit viral replication.

Hypothesis on the mechanisms supporting a molecule's antiviral effect

We cannot rule out that some in silico selected drugs exert a possible direct effect on a viral activity or on a cellular pathway exploited by the virus. Among the seven molecules, three in particular could have such an effect: ribavirin and merbromin which could both directly inhibit a viral function, and harmol which could inhibit a proviral pathway. Harmol is a beta-carboline alkaloid of the medicinal plant, *Perganum harmala L.* (Zygophylaceae). Few specific effects are described for harmol except that it exerts a psychoactive effect by inhibiting monoamine oxydase [57], moderately inhibits platelet aggregation by inhibiting PLCγ2 [58] and induces apoptosis in some cell lines by activating caspase 8 [59]. PLCγ2 is implicated in the protein kinase C (PKC) activation pathway, the activity of which is crucial for influenza virus entry [60]. Therefore its inhibition by harmol could in part be responsible for the antiviral effect shown by this molecule. Likewise, activation of apoptosis could limit viral replication [61].

However, three types of evidence support our hypothesis that the selected molecules have an antiviral effect by modifying the host cell gene expression. First, the results of our test of infection efficiencies demonstrate that none of the molecules except for merbromin had an effect on viral structure or function before infection (Figure 6). Second, the high confirmation rate of the in silico selected drug panel validate the rational of the selection. Last, some molecules that regulated the host cell transcription in the same way that influenza virus infection enhanced viral production.

To our knowledge, modulation of the cell gene expression has never been described to support the effects of the in silico selected drug, except for ribavirin. This antiviral drug with in vitro activity against both DNA and RNA viruses [62], has several mechanisms of action proposed to support its antiviral effect (reviewed in [63]: i) the depletion of the intracellular GTP-pool by inhibition of inosine monophosphate dehydrogenase compromises the synthesis of progeny viral RNA; ii) the inhibition of viral RNA-dependent RNA polymerase activity has been shown for hepatitis C and influenza viruses; and iii) it could act as a RNA virus mutagen causing error catastrophe). Which mechanisms contribute to its anti-influenza effect in vivo remains undetermined. In this study, we selected ribavirin because it inversed the gene expression signature of infection, which could highlight a new potential antiviral mechanism of this molecule. An effect of ribavirin on the cellular gene expression has been reported to contribute to its antiviral effect on the respiratory syncytial virus (RSV) [64] and the hepatitis C virus (HCV) [65]. In these studies, ribavirin enhanced the expression of *ISG* in infected cells. It was concluded that ribavirin potentiates the interferon response induced by peginterferon (during treatment of HCV) [65] or induced by RSV infection [64]. However, ribavirin has also been shown to alter the expression of many genes implicated in various other cellular pathways such as apoptosis [66], cell cycle control or intracellular signaling [64]. We propose that these modifications contribute to its antiviral effect.

Does this study now allow us to define co-factors and antiviral proteins?

None of the selected molecules fully inversed the infection signature. Therefore to try to identify anti or proviral factors, we first searched for genes whose expression could be inverted by all effective molecules. This was the case for only one gene, *calpain 1*, which was up-regulated by all the selected molecules and down-regulated during infection. The calpains, or calcium-regulated non-lysosomal thiol-proteases, are ubiquitous enzymes which catalyze limited proteolysis of substrates involved in cytoskeletal remodeling and signal transduction. We found no data in the literature describing any antiviral role for calpain 1. Such potential activity remains to be tested in the future.

It is also possible that each different molecule exerts its antiviral effect through different mechanisms and different combinations of gene expression modifications could be implied. These changes are listed in the Connectivity Map but except for midodrine and ribavirin, have yet to be confirmed by other studies. Midodrine is the prodrug of desglymidodrine, which is an alpha1-adrenergic receptor agonist used in the clinical management of patients with orthostatic hypotension [67]. Its effect on cellular gene expression can be derived from several microarray studies [68,69] showing many transcriptional changes after stimulation of the alpha1 adrenoreceptor, involving for example genes encoding integrin-mediated cell adhesion proteins and proteins involved in hyaluronan signaling [70]. These observations are consistent with the observed midodrine-induced downregulation of *ICAM1* and *HYAL4* reported in the Connectivity Map. Both of these genes were up-regulated during infection. Their potential role in the influenza cell cycle remains to be determined.

Recently, several human RNAi screens identified host cell factors which are required for influenza virus replication [37,71,72,73]. We wondered if the 20 genes of the concise infection signature were found to be important for the influenza virus in any of these screens (Table S5). Notably, the concise infection signature is specifically more enriched in regulators of influenza infection than random chance (compared to 8676 genes of the array, p-value = 0.0072 with Fisher's exact test). Four genes (*SLC2A2*, *ICAM1*, *OAS1*, *ISG15*) out of the 12 up-regulated genes were defined as proviral factors in these screens [37,71,72]. Three genes are ISGs: *ICAM1*, *OAS1* and *ISG15* that may be co-opted by the virus. Their down-regulation by the drugs could support partially their antiviral activity. On the other hand, none antiviral factor was identified in the list of 8 genes down-regulated during infection. This could be due to the low number of antiviral factors found by published screens (180 in total [37,71], compared to 875 (unique) proviral factors [37,71,72,73]). Therefore, the down-regulated genes of the infection signature can be considered as potential antiviral factors, which should be further tested.

Outcomes and perspectives

To conclude, our investigation of transcriptional profiles of cells infected with different strains of influenza A viruses highlights virus specificity but, above all, has allowed us to define a universal influenza A virus-induced gene expression signature. Here we

proposed to correlate this signature to gene expression profiles of cells treated by different molecules. This is the first study using the Connectivity Map to identify antivirals thought to act at the genomic level. One considerable advantage of some of these antivirals is their potential broad spectrum of action against all influenza A viruses, including novel pandemic viruses such as the H1N1 SOIV.

Except for harmol, all the antiviral molecules tested in this assay are approved for various different therapeutic indications. Our drug repositioning strategy should therefore contribute to the discovery of new alternative antivirals with accelerated regulatory registration. In the event of an unknown emerging virus, this approach may be of great interest to relatively quickly identify all available commercialized drugs with potential antiviral effects.

This study conducted in vitro in an established human cell line and with a nylon microarray represents a first proof of principle study. To identify effective anti-influenza molecules for use in clinical practice, we will now study the transcriptional response of patients to infection using a pan-genome microarray. Gene response to infection within a tissue in vivo should be more complex, with many cell types being implicated and with those infected being influenced by cytokines and the surrounding tissue.

Importantly, our dual experimental approach associating transcriptional study and in silico screening could be transferred to other pathogens. We are interested in identifying common gene-expression signatures of different viruses causing the same clinical disease to find useful therapies before etiologic diagnosis.

Supporting Information

Figure S1 Boxplots for the 7 interferon stimulated genes (ISGs) upregulated during infection. These boxplots show the averaged normalized expression values of the 7 ISGs in each group of samples. The bottom edge of the box represents the 25th percentile of the data while the top edge of the boxplot represents the 75th percentile. The line inside the box represents the 50th percentile of the data or the median.
Found at: doi:10.1371/journal.pone.0013169.s001 (0.13 MB TIF)

Figure S2 Neuraminidase assay is suitable for evaluation of the drug panel. A. Neuraminidase activity can be used to quantify virus. Different influenza viral stocks (A/New Caledonia/20/99 (H1N1), A/Moscow/10/99 (H3N2), A/Lyon/969/09 (H1N1 SOIV), A/Turkey/582/2006 (H5N1), A/Finch/England/2051/94 (H5N2), and A/Chicken/Italy/2076/99 (H7N1)) were serially diluted in DMEM. 25 µl of the viral dilutions were incubated with 75 µL of 20 µM MU-NANA for 1 hour at 37°C. Stop solution was added before reading the fluorescence. The signal to background ratio at each TCID50 is shown. Under these conditions, the signal to background ratio was proportional to the amount of virus between 2 and 70. At hightest titer, the enzyme activity reaches a plateau due to limiting substrate. Triton X-100 treatment of H5N1 viral stock enhanced NA activities, as previously described [27]. B. Molecules did not inhibit the neuraminidase activity of the influenza virus. A/Moscow/10/99 (H3N2) viral stock diluted in DMEM (107,8 TICD50/mL final) was incubated with increasing concentration of molecule (or DMEM for controls) for 0.5 h at room temperature before testing the neuraminidase activity as described in materials and methods.
Found at: doi:10.1371/journal.pone.0013169.s002 (0.52 MB TIF)

Figure S3 Effective molecules inhibit H3N2 viral growth at early and later times of infection. A. A549 cells were treated with increasing concentrations of the molecule or solvent for 6 h and were subsequently infected with A/Moscow/10/99 (H3N2) influenza virus at a moi of 2. A neuraminidase test was performed at 24, 42 and 65 hpi to assess influenza viral growth. Values represent the mean of two independent experiments performed in duplicate. B. Potency of the inhibitors against A/Moscow/10/99 (H3N2) according to different times of infection. CC50: molecule concentration of 50% cytotoxicity; EC50: molecule concentration of 50% inhibition of viral replication; SI: selective index.
Found at: doi:10.1371/journal.pone.0013169.s003 (0.89 MB TIF)

Figure S4 Concentration-viability curves of the eight molecules. A549 cells were treated with increasing concentrations of each molecule or solvent for 72 h and their viability was measured using Neutral Red dye (as described in materials and methods). Data are presented as a ratio of absorbance at 550 nm of treated cells to control cells. Values represent the mean of six independent experiments performed in duplicate, and error bars show the standard deviation of the mean. (+/− standard deviation). Horizontal lines are drawn to show the scatter of the control values (mean +/− standard deviation).
Found at: doi:10.1371/journal.pone.0013169.s004 (0.34 MB TIF)

Figure S5 Six molecules inhibited H3N2 influenza viral growth. A549 cells were treated with increasing concentrations of the molecule or solvent for 6 h and were subsequently infected with A/Moscow/10/99 (H3N2) influenza A virus at a moi of 0.2 (Panel A) or 2 (Panel B). Viral titers were determined at 65 hpi using a neuraminidase test as described in materials and methods. Results in A and B are representative of three independent determinations in duplicate. Data are presented as a ratio of RFU (relative fluorescence unit) in supernatants of treated cells to RFU in control cells (mean +/− standard deviation). The dose-response curves are the results of a fit with a 3-parameter logistic equation (if at least one normalized response was less than 0.5). Horizontal lines are drawn to show the scatter of the control values (mean +/− standard deviation).
Found at: doi:10.1371/journal.pone.0013169.s005 (0.68 MB TIF)

Figure S6 Concentration-viability curves of the seven positively correlated molecules. A549 cells were treated with increasing concentrations of each molecule or solvent for 72 h and their viability was measured using Neutral Red dye (as described in materials and methods). Data are presented as a ratio of absorbance at 550 nm of treated cells to control cells. Values represent the mean of three independent experiments performed in duplicate, and error bars show the standard deviation of the mean. (+/− standard deviation). Horizontal lines are drawn to show the scatter of the control values (mean +/− standard deviation).
Found at: doi:10.1371/journal.pone.0013169.s006 (0.30 MB TIF)

Figure S7 Three molecules enhanced H3N2 influenza viral growth at a moi of 0.2. A549 cells were treated with increasing concentrations of the molecule or solvent for 6 h and were subsequently infected with A/Moscow/10/99 (H3N2) influenza A virus at a moi of 0.2 (Panel A) or 2 (Panel B). Viral titers were determined at 65 hpi using a neuraminidase test as described in materials and methods. Results in A and B are representative of two independent determinations in duplicate. Data are presented as a ratio of RFU (relative fluorescence unit) in supernatants of treated cells to RFU in control cells (mean +/− standard deviation). The dose-response curves are the results of a fit with a 3-parameter logistic equation (if at least one normalized response was less than 0.5). Horizontal lines are drawn to show the scatter of the control values (mean +/− standard deviation). Enhancement of H3N2 virus replication was verified by measuring viral titers at 65 hpi by end point titration assays in MDCK cells (TCID50/mL).

For these assays, sample supernatants of duplicate were harvested at 65 hpi and stored at −80°C until analysis. Increase of viral titers were statistically significant for alvespimycin, methylbenzethoniumchloride, and sulodictil 40 μM (p-value <0.05, two tailed Welch t-test).
Found at: doi:10.1371/journal.pone.0013169.s007 (0.64 MB TIF)

Figure S8 Five molecules inhibit H1N1 SOIV influenza viral growth. A549 cells were treated with increasing concentrations of the molecule or solvent for 6 h and were subsequently infected with A/Lyon/969/09 (H1N1 SOIV) influenza virus at a moi of 0.2 (Panel A) or 2 (Panel B). Viral titers were determined at 65 hpi using a neuraminidase test as described in materials and methods. This assay was performed in duplicate. Data are presented as the ratio of RFU in the supernatant of treated cells to RFU in control cells (mean +/− standard deviation). The dose-response curves are the results of a fit with 3-parameter logistic equation (if at least one normalized response was less than 0.5). Horizontal lines are drawn to show the scatter of the control values (mean +/− standard deviation).
Found at: doi:10.1371/journal.pone.0013169.s008 (0.69 MB TIF)

Figure S9 Comparison of viral replication kinetics between different influenza A viruses in A549 cells. A549 cells were infected at a moi of 0.1 (Panel A) or at a moi of 2 (Panel B) with influenza virus A/New Caledonia/20/99 (H1N1), A/Lyon/969/09 (H1N1 SOIV), A/Moscow/10/99 (H3N2), A/Turkey/582/2006 (H5N1), A/Finch/England/2051/94 (H5N2), and A/Chicken/Italy/2076/99 (H7N1).
Found at: doi:10.1371/journal.pone.0013169.s009 (0.45 MB TIF)

Table S1 Specific primers used for real-time quantitative RT-PCR.
Found at: doi:10.1371/journal.pone.0013169.s010 (0.01 MB XLS)

Table S2 Gene expression data for 300 influenza virus regulated genes identified in our experiments. Values are log2 normalized gene expression intensities (see Experimental Procedures).
Found at: doi:10.1371/journal.pone.0013169.s011 (0.28 MB XLS)

Table S3 Approved indications of the eight potential antivirals.
Found at: doi:10.1371/journal.pone.0013169.s012 (0.01 MB XLS)

Table S4 Potency of the positively correlated drugs against A/Moscow/10/99 (H3N2). CC50: molecule concentration of 50% cytotoxicity; EC50: molecule concentration of 50% inhibition of viral replication; SI: selective index.
Found at: doi:10.1371/journal.pone.0013169.s013 (0.01 MB XLS)

Table S5 Comparison between the 300 regulated genes in our experiment and identified factors required for influenza replication in published human RNAi screens. Genes belonging to the concise signature used to query connectivity map in our study are highlighted in grey. Negative regulators of influenza replication (antiviral) are genes whose depletion enhance viral infection. Positive regulators (proviral) knock-out decrease viral infection. Antiviral factors were identified in Shapira et al and Brass et al studies.
Found at: doi:10.1371/journal.pone.0013169.s014 (0.03 MB XLS)

Acknowledgments

We thank Gaelle Cartet for technical assistance and Matthieu Yver, Dr Vincent Moules and Dr Emilie Frobert for their help in the BSL-3 laboratory. We are grateful to Dr Joel Lachuer (Profilexpert) for critical reading of the manuscript.

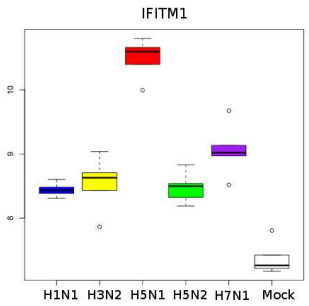
Author Contributions

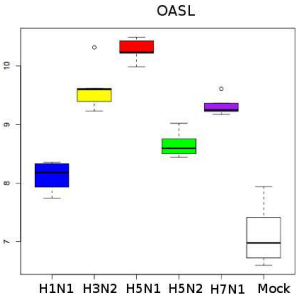
Conceived and designed the experiments: LJ JT JJD MRC. Performed the experiments: LJ JT BL. Analyzed the data: LJ JT JJD MRC. Contributed reagents/materials/analysis tools: LJ JT OF BL CN JJD MRC. Wrote the paper: LJ JT JJD MRC. Prepared and manipulated viruses in BSL3 laboratory in order to perform experiments in answer to some of the queries of reviewers: VM.

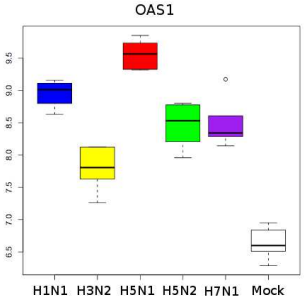
References

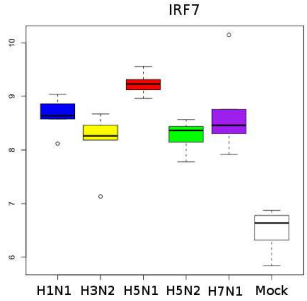
1. Kash JC (2009) Applications of high-throughput genomics to antiviral research: evasion of antiviral responses and activation of inflammation during fulminant RNA virus infection. Antiviral Research 83: 10–20.
2. Reeves JD, Piefer AJ (2005) Emerging drug targets for antiretroviral therapy. Drugs 65: 1747–1766.
3. Greco A, Callé A, Morfin F, Thouvenot D, Cayre M, et al. (2005) S-adenosyl methionine decarboxylase activity is required for the outcome of herpes simplex virus type 1 infection and represents a new potential therapeutic target. The FASEB Journal: 19: 1128–1130.
4. Ludwig S (2009) Targeting cell signalling pathways to fight the flu: towards a paradigm change in anti-influenza therapy. The Journal of Antimicrobial Chemotherapy 64: 1–4.
5. Lamb RA, Krug RM (2001) Orthomyxoviridae: the viruses and their replication. In: Howley DMKaPM, ed. Fields's Virology. Philadelphia: Lippincott Williams and Wilkins. pp 1487–1531.
6. Michaelis M, Doerr H, Cinatl J (2009) An Influenza A H1N1 Virus Revival - Pandemic H1N1/09 Virus. Infection 37: 381–389.
7. WHO (2009) WHO | Influenza (Seasonal). Available: http://www.who.int/mediacentre/factsheets/fs211/en. Accessed: 2009 December 21.
8. Hayden FG (2006) Antivirals for influenza: historical perspectives and lessons learned. Antiviral Research 71: 372–378.
9. Kash JC, Basler CF, García-Sastre A, Carter V, Billharz R, et al. (2004) Global host immune response: pathogenesis and transcriptional profiling of type A influenza viruses expressing the hemagglutinin and neuraminidase genes from the 1918 pandemic virus. Journal of Virology 78: 9499–9511.
10. Kash JC, Tumpey TM, Proll SC, Carter V, Perwitasari O, et al. (2006) Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus. Nature 443: 578–581.
11. Cameron CM, Cameron MJ, Bermejo-Martin JF, Ran L, Xu L, et al. (2008) Gene expression analysis of host innate immune responses during Lethal H5N1 infection in ferrets. Journal of Virology 82: 11308–11317.
12. Baskin CR, Bielefeldt-Ohmann H, Tumpey TM, Sabourin PJ, Long JP, et al. (2009) Early and sustained innate immune response defines pathology and death in nonhuman primates infected by highly pathogenic influenza virus. Proceedings of the National Academy of Sciences of the United States of America 106: 3455–3460.
13. Lamb J (2007) The Connectivity Map: a new tool for biomedical research. Nature Reviews Cancer 7: 54–60.
14. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313: 1929–1935.
15. Moules V, Ferraris O, Terrier O, Giudice E, Yver M, et al. (2010) In vitro characterization of naturally occurring influenza H3NA- viruses lacking the NA gene segment: Toward a new mechanism of viral resistance? Virology 404: 215–224.
16. Puthier D, Joly F, Irla M, Saade M, Victorero G, et al. (2004) A general survey of thymocyte differentiation by transcriptional analysis of knockout mouse models. Journal of Immunology 173: 6109–6118.
17. Lopez F, Rougemont J, Loriod B, Bourgeois A, Loï L, et al. (2004) Feature extraction and signal processing for nylon DNA microarrays. BMC Genomics 5: 38.
18. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 5: R80.
19. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America 98: 5116–5121.
20. Schwender H, Ickstadt K (2008) Empirical Bayes analysis of single nucleotide polymorphisms. BMC Bioinformatics 9: 144.
21. Benjamini Y, Hochberg, Y: (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 57: 289–300.

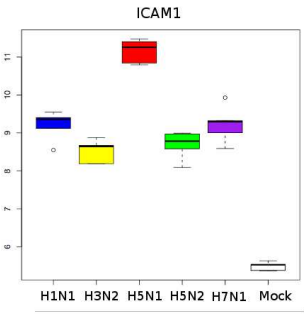
22. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402–408.
23. Fautz R, Husein B, Hechenberger C (1991) Application of the neutral red assay (NR assay) to monolayer cultures of primary hepatocytes: rapid colorimetric viability determination for the unscheduled DNA synthesis test (UDS). Mutation Research 253: 173–179.
24. Smee DF, Morrison AC, Barnard DL, Sidwell RW (2002) Comparison of colorimetric, fluorometric, and visual methods for determining anti-influenza (H1N1 and H3N2) virus activities and toxicities of compounds. Journal of Virological Methods 106: 71–79.
25. Fotakis G, Timbrell JA (2006) In vitro cytotoxicity assays: comparison of LDH, neutral red, MTT and protein assay in hepatoma cell lines following exposure to cadmium chloride. Toxicology Letters 160: 171–177.
26. Eichelberger MC, Hassantoufighi A, Wu M, Li M (2008) Neuraminidase activity provides a practical read-out for a high throughput influenza antiviral screening assay. Virology Journal 5: 109.
27. Jonges M, Liu WM, van der Vries E, Jacobi R, Pronk I, et al. (2010) Influenza Virus Inactivation for Studies of Antigenicity and Phenotypic Neuraminidase Inhibitor Resistance Profiling. J Clin Microbiol 48: 928–940.
28. DeLean A, Munson PJ, Rodbard D (1978) Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. The American Journal of Physiology 235: E97–102.
29. Ritz C, Streibig JC (2005) Bioassay Analysis using R. J Statist Software 12.
30. Guo Y, Rumschlag-Booms E, Wang J, Xiao H, Yu J, et al. (2009) Analysis of hemagglutinin-mediated entry tropism of H5N1 avian influenza. Virology Journal 6: 39.
31. Kanai Y, Chittaganpitch M, Nakamura I, Li G-M, Bai G-R, et al. (2010) Distinct propagation efficiencies of H5N1 influenza virus Thai isolates in newly established murine respiratory region-derived cell clones. Virus Research In Press.
32. Li IWS, Chan KH, To KWK, Wong SSY, Ho PL, et al. (2009) Differential susceptibility of different cell lines to swine-origin influenza A H1N1, seasonal human influenza A H1N1, and avian influenza A H5N1 viruses. Journal of Clinical Virology 46: 325–330.
33. Lee D, Mok C, Law A, Peiris M, Lau A (2010) Differential replication of avian influenza H9N2 viruses in human alveolar epithelial A549 cells. Virology Journal 7: 71.
34. Scull MA, Gillim-Ross L, Santos C, Roberts KL, Bordonali E, et al. (2009) Avian Influenza Virus Glycoproteins Restrict Virus Replication and Spread through Human Airway Epithelium at Temperatures of the Proximal Airways. PLoS Pathog 5: e1000424.
35. Greve JM, Davis G, Meyer AM, Forte CP, Yost SC, et al. (1989) The major human rhinovirus receptor is ICAM-1. Cell 56: 839–847.
36. Ning S, Hahn AM, Huye LE, Pagano JS (2003) Interferon regulatory factor 7 regulates expression of Epstein-Barr virus latent membrane protein 1: a regulatory circuit. Journal of Virology 77: 9359–9368.
37. Shapira SD, Gat-Viks I, Shum BOV, Dricot A, de Grace MM, et al. (2009) A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. Cell 139: 1255–1267.
38. Baskin CR, García-Sastre A, Tumpey TM, Bielefeldt-Ohmann H, Carter VS, et al. (2004) Integration of clinical data, pathology, and cDNA microarrays in influenza virus-infected pigtailed macaques (Macaca nemestrina). Journal of Virology 78: 10420–10432.
39. Tong HH, Long JP, Li D, DeMaria TF (2004) Alteration of gene expression in human middle ear epithelial cells induced by influenza A virus and its implication for the pathogenesis of otitis media. Microbial Pathogenesis 37: 193–204.
40. Schmolke M, Viemann D, Roth J, Ludwig S (2009) Essential impact of NF-kappaB signaling on the H5N1 influenza A virus-induced transcriptome. Journal of Immunology 183: 5180–5189.
41. Baas T, Baskin CR, Diamond DL, García-Sastre A, Bielefeldt-Ohmann H, et al. (2006) Integrated molecular signature of disease: analysis of influenza virus-infected macaques through functional genomics and proteomics. Journal of Virology 80: 10813–10828.
42. Smee DF, Huffman JH, Morrison AC, Barnard DL, Sidwell RW (2001) Cyclopentane neuraminidase inhibitors with potent in vitro anti-influenza virus activities. Antimicrobial Agents and Chemotherapy 45: 743–748.
43. Hoffmann HH, Palese P, Shaw ML (2008) Modulation of influenza virus replication by alteration of sodium ion transport and protein kinase C activity. Antiviral Research 80: 124–134.
44. Klein M, Brewer JH (1948) The inactivation of influenza virus by mercurials and the reactivation of sodium thioglycolate and BAL. Journal of Immunology 59: 135–140.
45. Sidwell RW, Bailey KW, Wong MH, Barnard DL, Smee DF (2005) In vitro and in vivo influenza virus-inhibitory effects of viramidine. Antiviral Research 68: 10–17.
46. Ludwig S, Planz O, Pleschka S, Wolff T (2003) Influenza-virus-induced signaling cascades: targets for antiviral therapy? Trends in Molecular Medicine 9: 46–52.
47. Naffakh N, Tomoiu A, Rameix-Welti M-A, van der Werf S (2008) Host Restriction of Avian Influenza Viruses at the Level of the Ribonucleoproteins. Annual Review of Microbiology 62: 403–424.
48. de Jong MD, Simmons CP, Thanh TT, Hien VM, Smith GJD, et al. (2006) Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. Nature Medicine 12: 1203–1207.
49. Chan RWY, Yuen KM, Yu WCL, Ho CCC, Nicholls JM, et al. (2010) Influenza H5N1 and H1N1 Virus Replication and Innate Immune Responses in Bronchial Epithelial Cells Are Influenced by the State of Differentiation. PLoS One 5(1): e8713.
50. Marjuki H, Yen H-L, Franks J, Webster RG, Pleschka S, et al. (2007) Higher polymerase activity of a human influenza virus enhances activation of the hemagglutinin-induced Raf/MEK/ERK signal cascade. Virology Journal 4: 134.
51. Geiss GK, An MC, Bumgarner RE, Hammersmark E, Cunningham D, et al. (2001) Global impact of influenza virus on cellular pathways is mediated by both replication-dependent and -independent events. Journal of Virology 75: 4321–4331.
52. Katze MG, Krug RM (1984) Metabolism and expression of RNA polymerase II transcripts in influenza virus-infected cells. Molecular and Cellular Biology 4: 2198–2206.
53. Hale BG, Randall RE, Ortín J, Jackson D (2008) The multifunctional NS1 protein of influenza A viruses. The Journal of General Virology 89: 2359–2376.
54. Bjornson AB, Mellencamp MA, Schiff GM (1991) Complement is activated in the upper respiratory tract during influenza virus infection. The American Review of Respiratory Disease 143: 1062–1066.
55. Zhang H, Su YA, Hu P, Yang J, Zheng B, et al. (2006) Signature patterns revealed by microarray analyses of mice infected with influenza virus A and Streptococcus pneumoniae. Microbes and Infection 8: 2172–2185.
56. El Kasmi KC, Smith AM, Williams L, Neale G, Panopolous A, et al. (2007) Cutting Edge: A Transcriptional Repressor and Corepressor Induced by the STAT3-Regulated Anti-Inflammatory Signaling Pathway. J Immunol 179: 7215–7219.
57. Herraiz T, González D, Ancín-Azpilicueta C, Arán VJ, Guillén H (2009) beta-Carboline alkaloids in Peganum harmala and inhibition of human monoamine oxidase (MAO) Food and Chemical Toxicology 48: 839–345.
58. Im J-H, Jin Y-R, Lee J-J, Yu J-Y, Han X-H, et al. (2009) Antiplatelet activity of beta-carboline alkaloids from Perganum harmala: a possible mechanism through inhibiting PLCgamma2 phosphorylation. Vascular Pharmacology 50: 147–152.
59. Abe A, Yamada H (2009) Harmol induces apoptosis by caspase-8 activation independently of Fas/Fas ligand interaction in human lung carcinoma H596 cells. Anti-Cancer Drugs 20: 373–381.
60. Root CN, Wills EG, McNair LL, Whittaker GR (2000) Entry of influenza viruses into cells is inhibited by a highly specific protein kinase C inhibitor. The Journal of General Virology 81: 2697–2705.
61. Kurokawa M, Koyama AH, Yasuoka S, Adachi A (1999) Influenza virus overcomes apoptosis by rapid multiplication. International Journal of Molecular Medicine 3: 527–530.
62. Chan-Tack KM, Murray JS, Birnkrant DB (2009) Use of ribavirin to treat influenza. The New England Journal of Medicine 361: 1713–1714.
63. Leyssen P, De Clercq E, Neyts J (2008) Molecular strategies to inhibit the replication of RNA viruses. Antiviral Research 78: 9–25.
64. Zhang Y, Jamaluddin M, Wang S, Tian B, Garofalo RP, et al. (2003) Ribavirin Treatment Up-Regulates Antiviral Gene Expression via the Interferon-Stimulated Response Element in Respiratory Syncytial Virus-Infected Epithelial Cells. Journal of Virology 77: 5933–5947.
65. Feld JJ, Nanda S, Huang Y, Chen W, Cam M, et al. (2007) Hepatic Gene Expression During Treatment with Peginterferon and Ribavirin: Identifying Molecular Pathways for Treatment Response. Hepatology 46: 1548–1563.
66. Feld JJ, Lutchman GA, Heller T, Hara K, Pfeiffer JK, et al. (2010) Ribavirin improves early responses to peginterferon through enhanced interferon signaling. Gastroenterology 139: 154–162.
67. Cruz DN (2000) Midodrine: a selective alpha-adrenergic agonist for orthostatic hypotension and dialysis hypotension. Expert Opinion on Pharmacotherapy 1: 835–840.
68. Gonzalez-Cabrera PJ, Gaivin RJ, Yun J, Ross SA, Papay RS, et al. (2003) Genetic Profiling of α1-Adrenergic Receptor Subtypes by Oligonucleotide Microarrays: Coupling to Interleukin-6 Secretion but Differences in STAT3 Phosphorylation and gp-130. Molecular Pharmacology 63: 1104–1116.
69. Gonzalez-Cabrera PJ, Shi T, Yun J, McCune DF, Rorabaugh BR, et al. (2004) Differential Regulation of the Cell Cycle by {alpha}1-Adrenergic Receptor Subtypes. Endocrinology 145: 5157–5167.
70. Shi T, Duan Z-H, Papay R, Pluskota E, Gaivin RJ, et al. (2006) Novel α1-Adrenergic Receptor Signaling Pathways: Secreted Factors and Interactions with the Extracellular Matrix. Molecular Pharmacology 70: 129–142.
71. Brass AL, Huang IC, Benita Y, John SP, Krishnan MN, et al. (2009) The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. Cell 139: 1243–1254.
72. Karlas A, Machuy N, Shin Y, Pleissner K-P, Artarini A, et al. (2010) Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. Nature 463: 818–822.
73. König R, Stertz S, Zhou Y, Inoue A, Heinrich Hoffmann H, et al. (2010) Human host factors required for influenza virus replication Nature 463: 813–817.

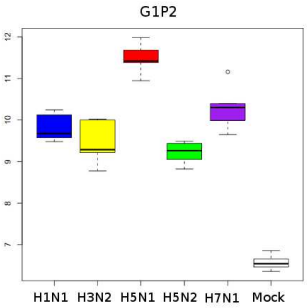

IFITM1
H1N1 H3N2 H5N1 H5N2 H7N1 Mock

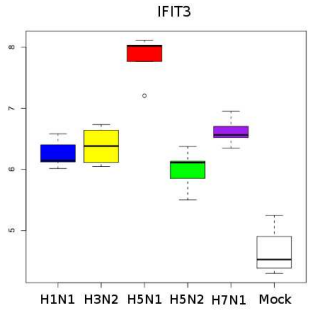

OASL
H1N1 H3N2 H5N1 H5N2 H7N1 Mock


OAS1
H1N1 H3N2 H5N1 H5N2 H7N1 Mock

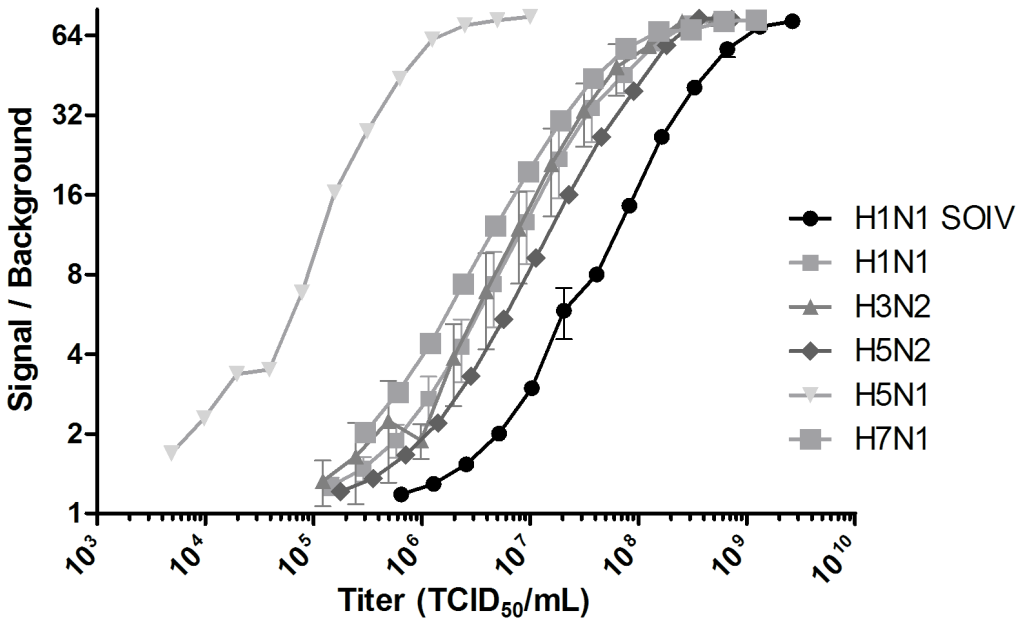

IRF7
H1N1 H3N2 H5N1 H5N2 H7N1 Mock


ICAM1
H1N1 H3N2 H5N1 H5N2 H7N1 Mock

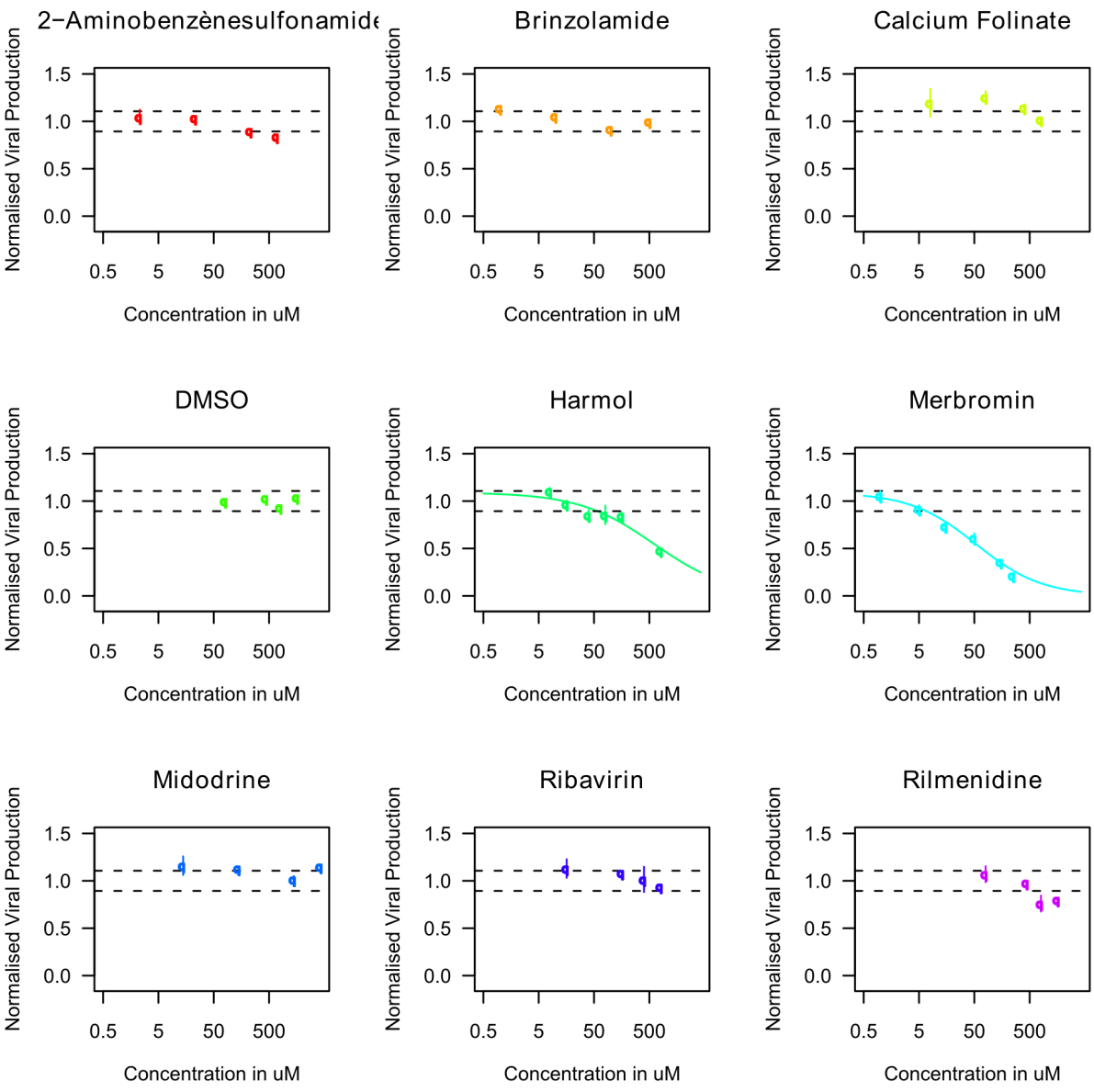

G1P2
H1N1 H3N2 H5N1 H5N2 H7N1 Mock


IFIT3
H1N1 H3N2 H5N1 H5N2 H7N1 Mock

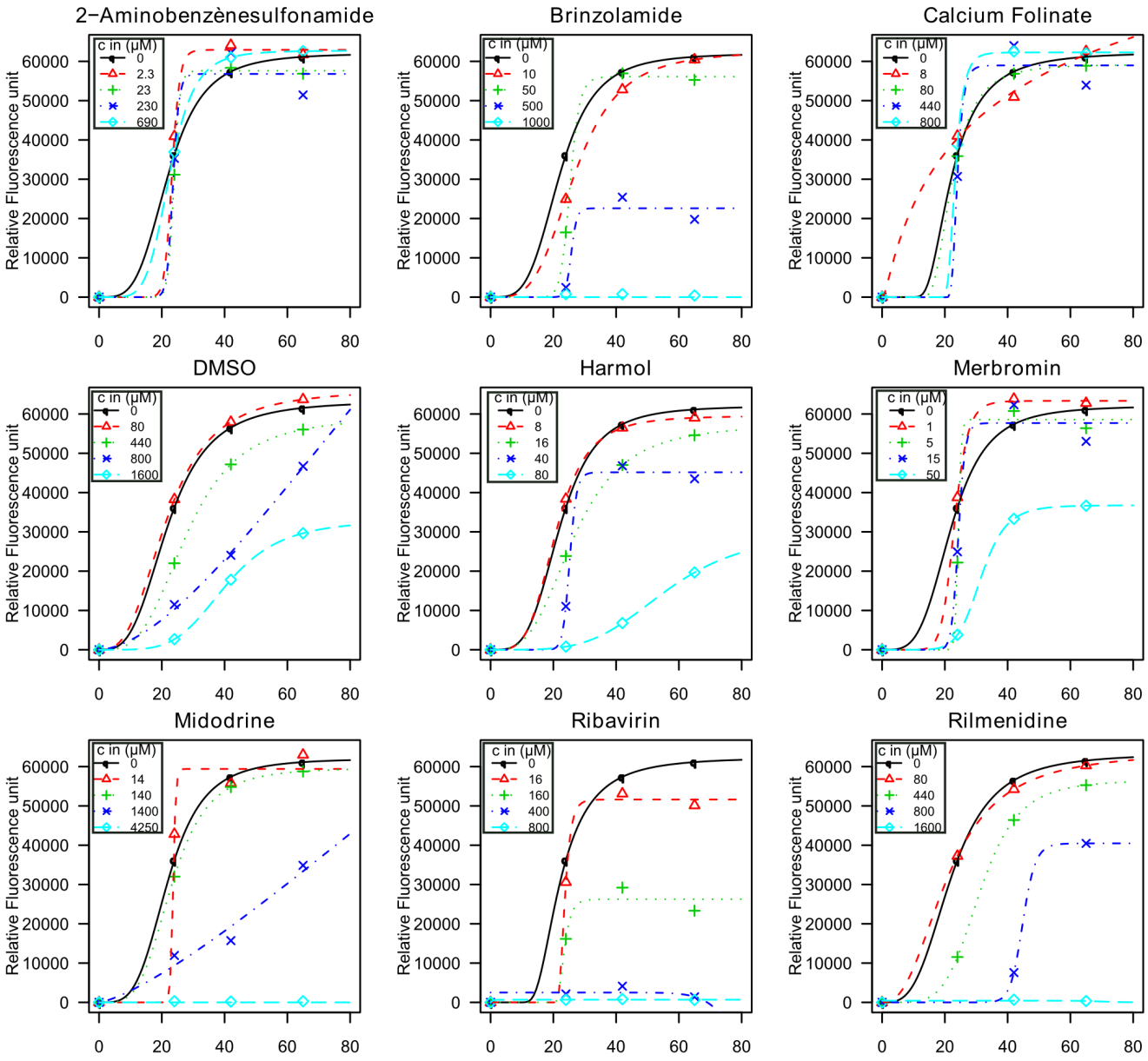
A


Signal / Background
64
32
16
8
4
2
1
10^3
10^4
10^5
10^6
10^7
10^8
10^9
10^{10}
Titer ($TCID_{50}$/mL)
H1N1 SOIV
H1N1
H3N2
H5N2
H5N1
H7N1

B

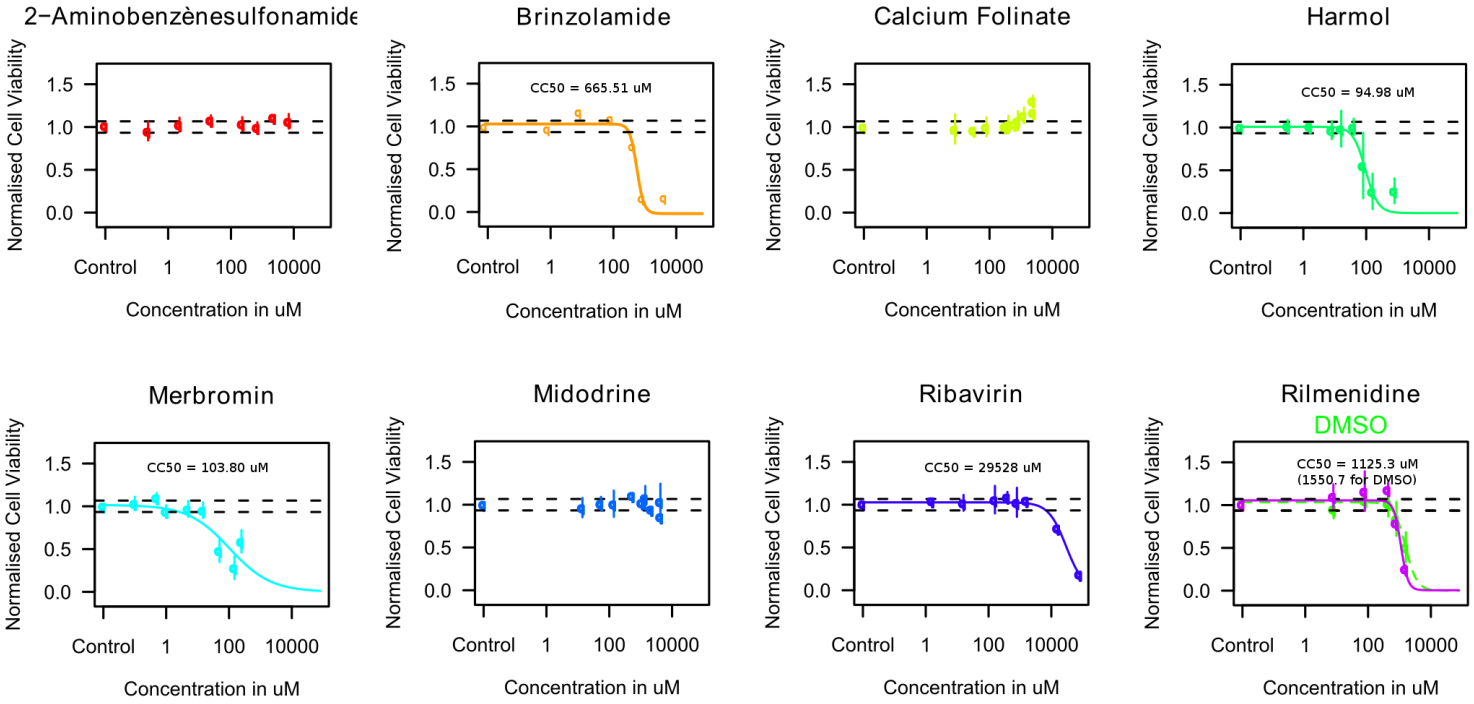

2−Aminobenzènesulfonamide
Brinzolamide
Calcium Folinate
DMSO
Harmol
Merbromin
Midodrine
Ribavirin
Rilmenidine
Normalised Viral Production
1.5
1.0
0.5
0.0
0.5
5
50
500
Concentration in uM

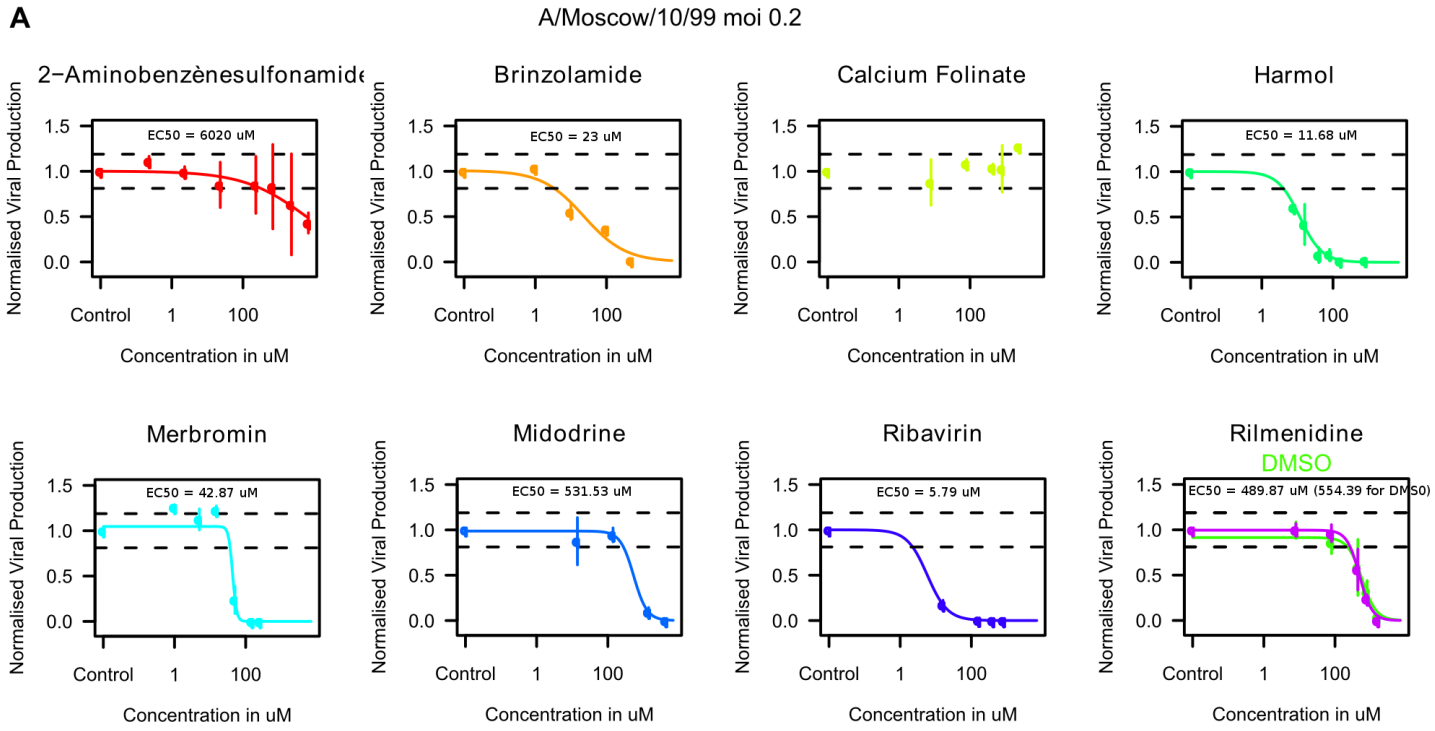
A

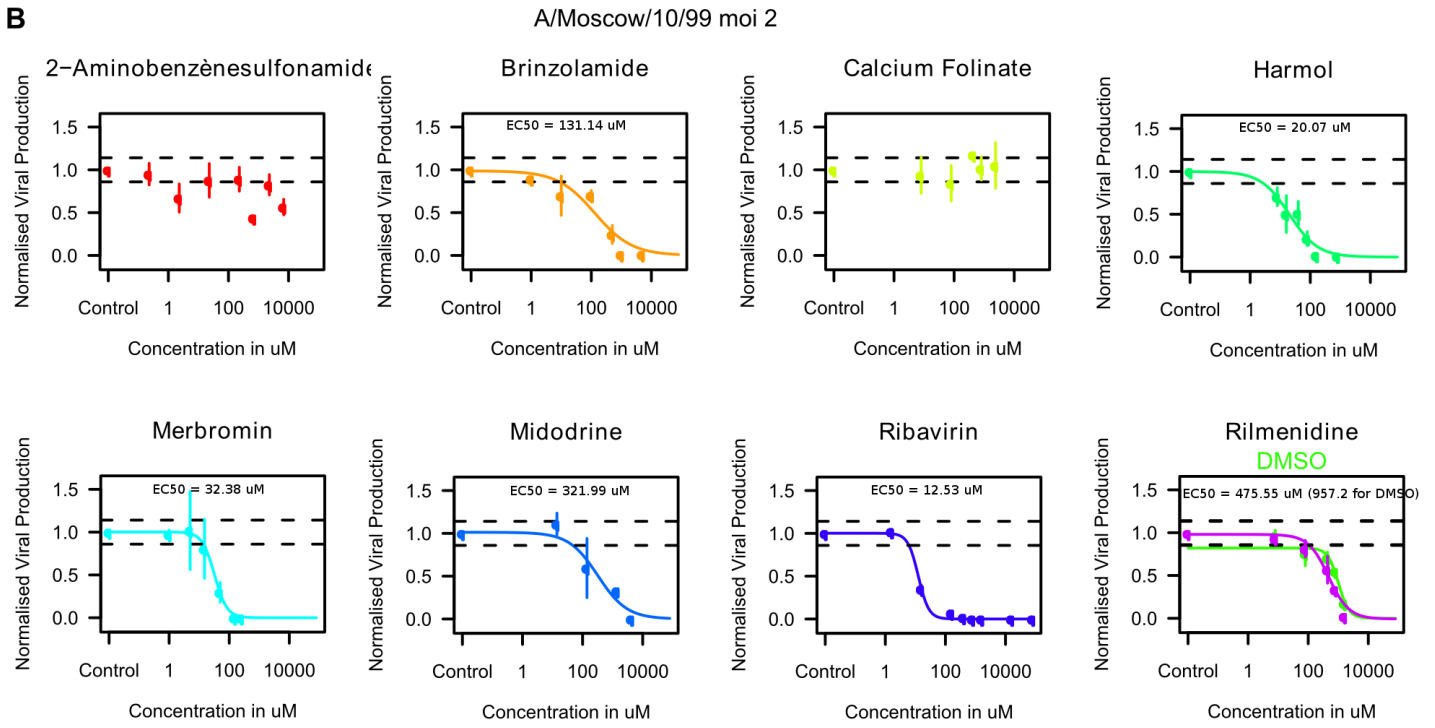


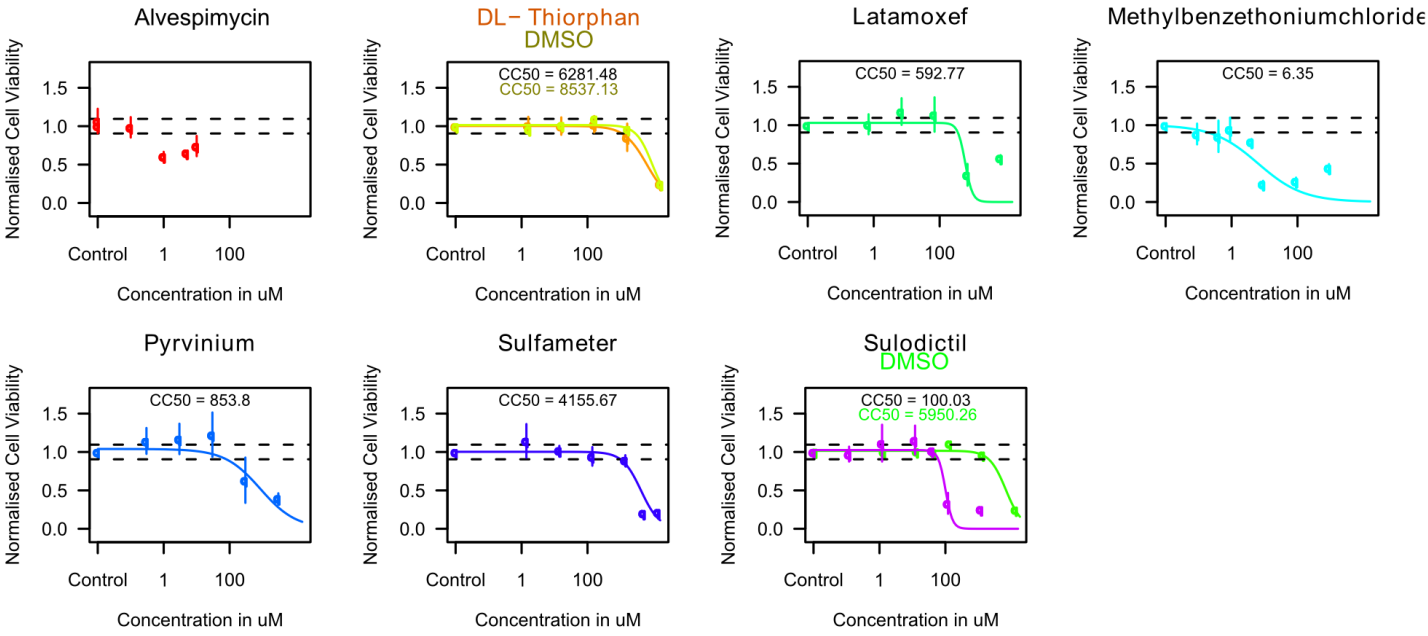
B

	24 hpi			42 hpi			65 hpi		
Substance	**CC50 (µM)**	**EC50 (µM)**	**SI**	**CC50 (µM)**	**EC50 (µM)**	**SI**	**CC50 (µM)**	**EC50 (µM)**	**SI**
2-Aminobenzenesulfonamide	> 6900	> 690	ND	> 6900	> 690	ND	> 6900	> 6900	ND
Brinzolamide	714,09	32,12	22,23	677,47	482,24	1,40	665,51	444,64	1,50
CalciµM Folinate	> 2400	> 800	ND	> 2400	> 800	ND	> 2400	> 2400	ND
Harmol	> 80	24,73	> 3,24	> 80	54,42	> 1,47	94,98	58,16	1,63
Merbromin	> 50	15,73	> 3,18	> 50	50,62	> 0,99	103,80	72,35	1,43
Midodrine	> 4250	736,12	> 5,77	> 4250	681,18	> 6,24	> 4250	1480,50	> 2,87
Ribavirin	53111,00	144,19	368,34	34540,00	162,65	212,36	29528,00	78,70	375,20
Rilmenidine (in DMSO)	> 1600	382,77 (546,53)	> 4,18 (2,93)	1478,69	577,51	2,56 (1,89)	1125,3 (1550,7)	897,36 (1519,4)	1,25 (1,02)

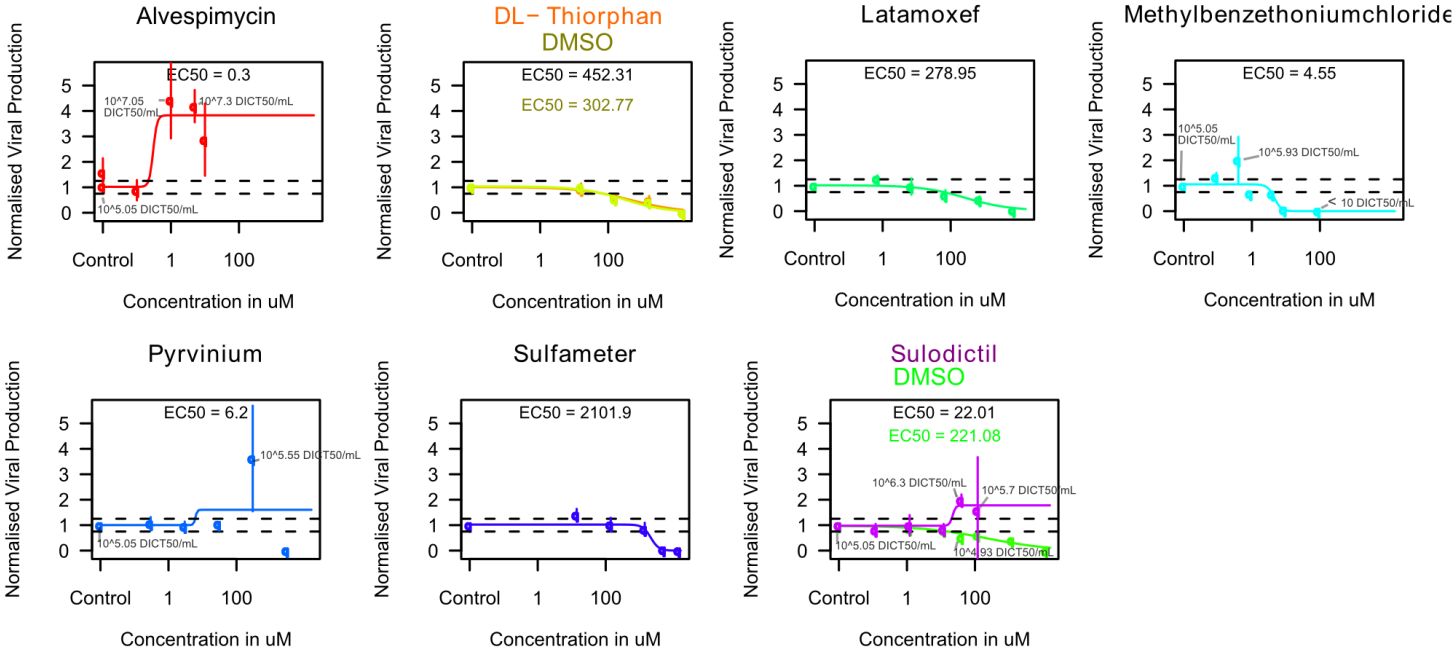

2−Aminobenzènesulfonamide
Brinzolamide
Calcium Folinate
Harmol
Merbromin
Midodrine
Ribavirin
Rilmenidine
DMSO
CC50 = 665.51 uM
CC50 = 94.98 uM
CC50 = 103.80 uM
CC50 = 29528 uM
CC50 = 1125.3 uM
(1550.7 for DMSO)
Normalised Cell Viability
Concentration in uM
Control
1
100
10000
0.0
0.5
1.0
1.5


A
A/Moscow/10/99 moi 0.2
2−Aminobenzènesulfonamide
EC50 = 6020 uM
Brinzolamide
EC50 = 23 uM
Calcium Folinate
Harmol
EC50 = 11.68 uM
Merbromin
EC50 = 42.87 uM
Midodrine
EC50 = 531.53 uM
Ribavirin
EC50 = 5.79 uM
Rilmenidine
DMSO
EC50 = 489.87 uM (554.39 for DMSO)
Normalised Viral Production
Control
1
100
Concentration in uM

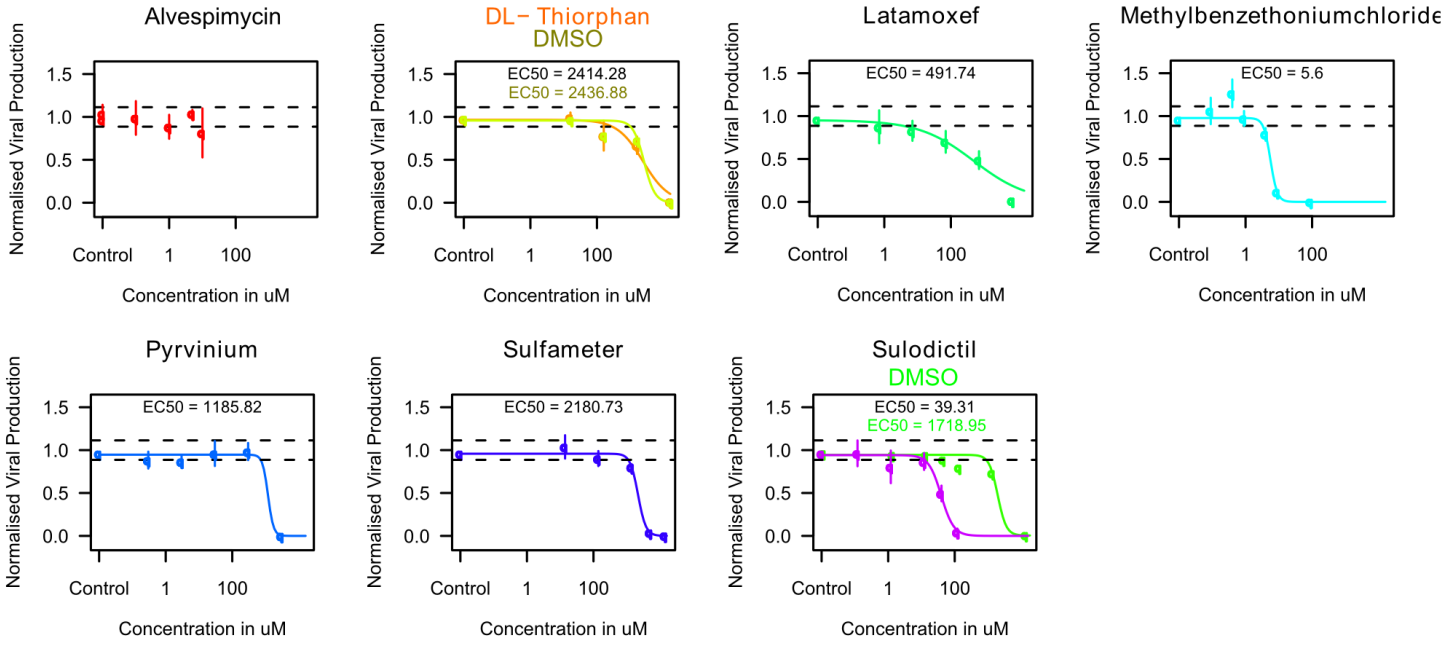

B
A/Moscow/10/99 moi 2
2−Aminobenzènesulfonamide
Brinzolamide
EC50 = 131.14 uM
Calcium Folinate
Harmol
EC50 = 20.07 uM
Merbromin
EC50 = 32.38 uM
Midodrine
EC50 = 321.99 uM
Ribavirin
EC50 = 12.53 uM
Rilmenidine
DMSO
EC50 = 475.55 uM (957.2 for DMSO)
Normalised Viral Production
Control
1
100
10000
Concentration in uM

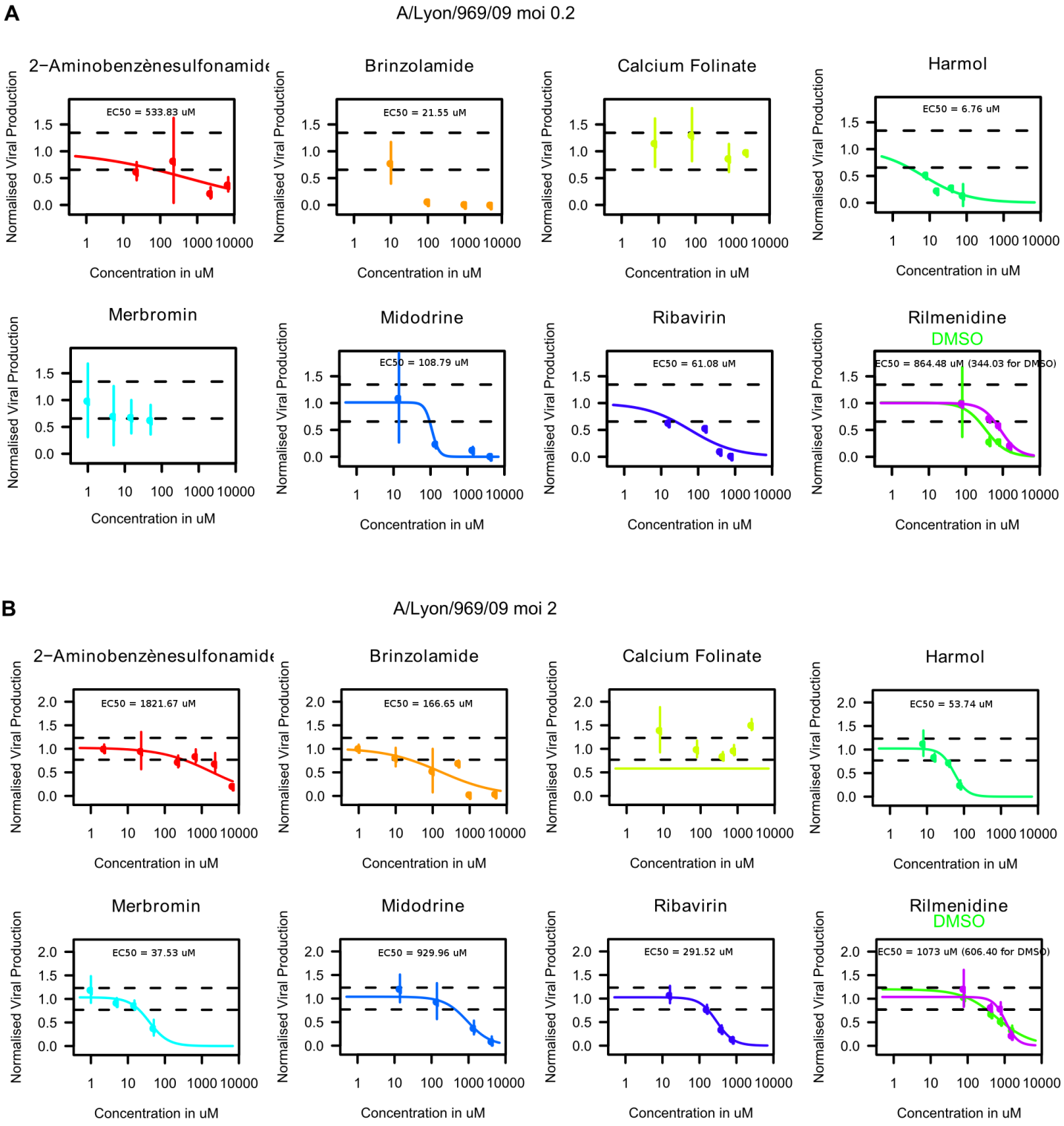

Alvespimycin
Normalised Cell Viability
1.5
1.0
0.5
0.0
Control
1
100
Concentration in uM
DL− Thiorphan
DMSO
CC50 = 6281.48
CC50 = 8537.13
Normalised Cell Viability
Concentration in uM
Latamoxef
CC50 = 592.77
Normalised Cell Viability
Concentration in uM
Methylbenzethoniumchloride
CC50 = 6.35
Normalised Cell Viability
Concentration in uM
Pyrvinium
CC50 = 853.8
Normalised Cell Viability
Concentration in uM
Sulfameter
CC50 = 4155.67
Normalised Cell Viability
Concentration in uM
Sulodictil
DMSO
CC50 = 100.03
CC50 = 5950.26
Normalised Cell Viability
Concentration in uM

A H3N2 moi 0.2

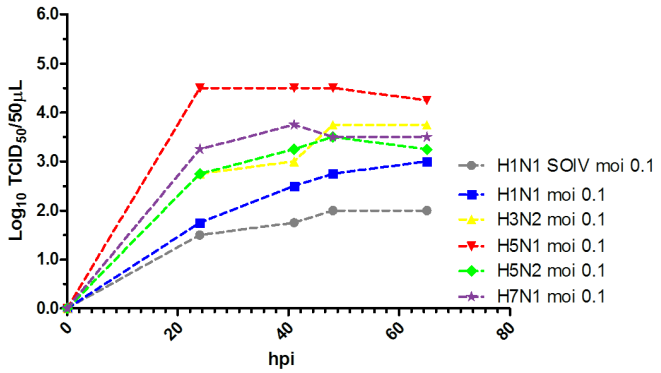


B H3N2 moi 2

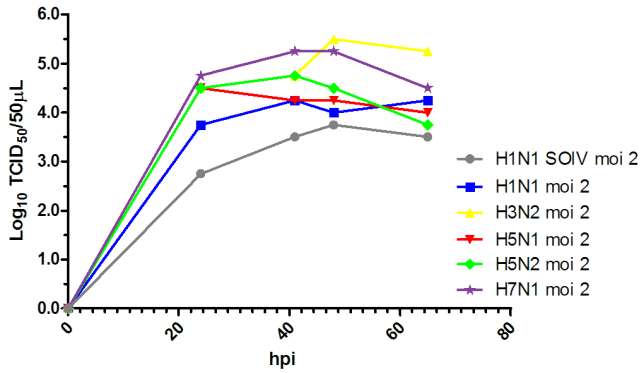



A
A/Lyon/969/09 moi 0.2
2−Aminobenzènesulfonamide
EC50 = 533.83 uM
Brinzolamide
EC50 = 21.55 uM
Calcium Folinate
Harmol
EC50 = 6.76 uM
Merbromin
Midodrine
EC50 = 108.79 uM
Ribavirin
EC50 = 61.08 uM
Rilmenidine
DMSO
EC50 = 864.48 uM (344.03 for DMSO)
Normalised Viral Production
Concentration in uM
B
A/Lyon/969/09 moi 2
2−Aminobenzènesulfonamide
EC50 = 1821.67 uM
Brinzolamide
EC50 = 166.65 uM
Calcium Folinate
Harmol
EC50 = 53.74 uM
Merbromin
EC50 = 37.53 uM
Midodrine
EC50 = 929.96 uM
Ribavirin
EC50 = 291.52 uM
Rilmenidine
DMSO
EC50 = 1073 uM (606.40 for DMSO)

A


6.0
5.0
4.0
3.0
2.0
1.0
0.0
Log10 TCID50/50µL
0
20
40
60
80
hpi
H1N1 SOIV moi 0.1
H1N1 moi 0.1
H3N2 moi 0.1
H5N1 moi 0.1
H5N2 moi 0.1
H7N1 moi 0.1

B


6.0
5.0
4.0
3.0
2.0
1.0
0.0
Log10 TCID50/50µL
0
20
40
60
80
hpi
H1N1 SOIV moi 2
H1N1 moi 2
H3N2 moi 2
H5N1 moi 2
H5N2 moi 2
H7N1 moi 2

Gene Symbol	Forward Primer	Reverse Primer
ICAM1	AGCTTCTCCTGCTCTGCAAC	AATCCCTCTCGTCCAGTCG
OAS1	CCAGGAAATTAGGAGACAGC	GAGCGAACTCAGTACGAAGC
G1P2	ACTCATCTTTGCCAGTACAGGAG	CAGCATCTTCACCGTCAGGTC
CAPN1	GCCAAAACATTGCCTGTTATCTTAG	ATAGGAGGCCGTATCAAAATTCC
NTE	GGCAACGTCATTGAGAAAATGC	AATCTCTGCCAAGTCAGTGAAGC
DNMT1	GTGGGGGACTGTGTCTCTGT	TGAAAGCTGCATGTCCTCAC
GAPD	GTGAAGGTCGGAGTCAACG	TGAGGTCAATGAAGGGGTC

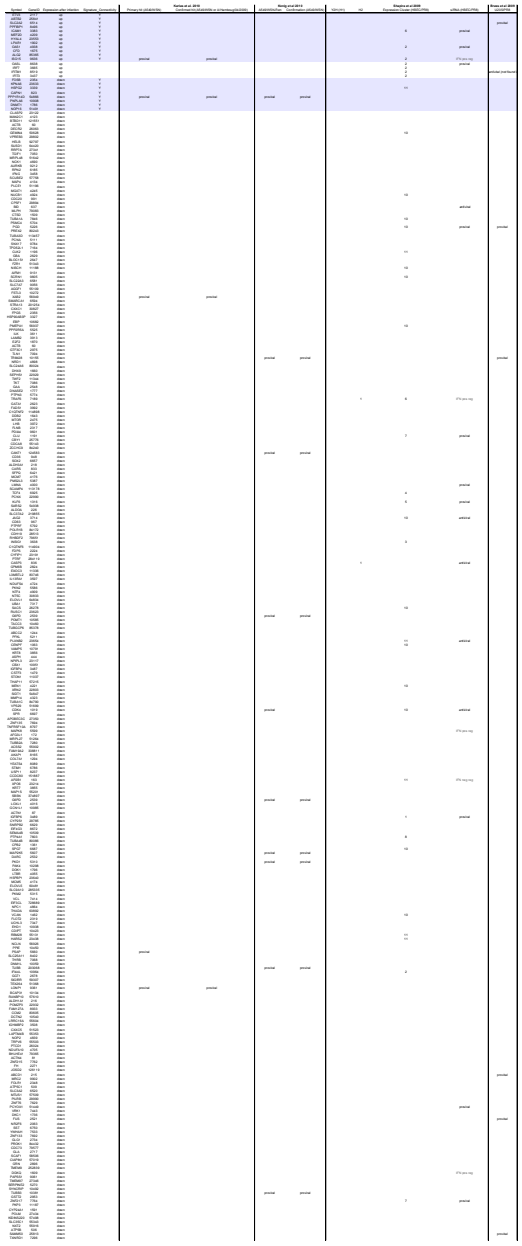
Table S2.

Gene expression data for 300 influenza virus regulated genes identified in our experiments. Values are log2 normalized gene expression intensities (see Experimental Procedures).

Ce fichier n'est pas reproduit compte tenu du nombre important de pages qu'il représente (tableau contenant les valeurs d'expression au sein de chaque échantillon des 300 gènes identifiés).

Name	Indications
2 aminobenzenesulfonamide	2 aminobenzenesulfonamide (or sulfanilamide) have antibacterial effects by inhibiting the production of folic acid inside the bacterial cell.
Brinzolamide	Brinzolamide is a carbonic anhydrase II (CA-II) inhibitor used topically to treat ocular hypertension and glaucoma
Calcium Folinate	Calcium folinate (or leucovirin) is a 5 formyl derivative of tetrahydrofolic acid which has vitamin activity equivalent to folic acid without requiring dihydrofolate reductase for its conversion. This compound is used as an adjuvant in cancer chemotherapy involving methotrexate.
Harmol	Harmol is one of the β-carboline indoleamine alkaloids in the plants *Peganum, Tribulus*, *Kallstroemia* spp. with various pharmacological activities and neuromodulator properties.
Merbromin	Merbromin (marketed as Mercurochrome) is a topical antiseptic used for minor cuts and scrapes.
Midodrine	Midodrine is an alpha1-adrenergic receptor agonist used in the clinical management of patients with orthostatic hypotension.
Ribavirin	Ribavirin is an antiviral drug with several proposed mechanisms. It is licensed for treatment of hepatitis C and respiratory syncytial virus.
Rilmenidine	Rilmenidine is an imidazoline derivative that appears to lower blood pressure by an interaction with imidazoline (I1) receptors in the brainstem (and kidneys).

		H3N2 moi 2		H3N2 moi 0.2	
Substance	CC50 (uM)	EC50 (uM)	SI = CC50/EC50	EC50 (uM)	SI = CC50/EC50
alvespimycin	> 10	> 10	ND	enhance viral production	ND
DL-Thiorphan (in DMSO)	6281.48 (8537.13)	2414.28 (2436.88)	2.60 (3.50)	452.31 (302.77)	13.89 (28.20)
Latamoxef	592,77	491,74	1,21	278,95	2,13
Methylbenzethoniumchloride	6,35	5,60	1,13	4,55	1,40
Pyrvinium	853,80	1185,82	0,72	enhance viral production	ND
Sulfameter	4155,67	2180,73	1,91	2101,90	1,98
Sulodictil (in DMSO)	100.03 (5950.26)	39.31 (1718.95)	2.54 (3.46)	enhance viral production	ND



Discussion

Discussion

La réponse immunitaire innée à l'encontre des micro-organismes est supposée être schématique et stéréotypée [77]. Sommairement, elle met en jeu une réponse humorale innée (anticorps naturels, défensines, système du complément) et un ensemble de cellules capables de réagir sans éducation préalable (lymphocytes NK, granulocytes, macrophages et cellules dendritiques). Les techniques d'analyse à haut débit montrent que cette simplification est excessive. La réponse de l'hôte à l'infection est basée sur des mécanismes différents en fonction de l'agent vulnérant et de l'hôte lui-même. L'arrivée d'outils tels que les puces à ADN permet par exemple d'explorer la relation hôte-pathogène avec infiniment plus de détails qu'auparavant [78].

Les interactions entre un hôte et un micro-organisme sont diverses et choréiformes. Les caractéristiques et le devenir de ces interactions diffèrent selon le type de cellule immunitaire recrutée, le sous-type de lymphocyte impliqué ou encore le profil de médiateurs produits. Il existe des thèmes communs dans la réponse de l'hôte à l'infection (par exemple les PAMPs pour « Pathogen Associated Molecular Patterns »). Parallèlement, il existe des thèmes communs dans les facteurs de virulence des micro-organismes (facteurs d'adhérence, toxines, systèmes de sécrétion, facteurs de régulation de l'expression du génome microbien) [79]. La description séparée de chaque partenaire (hôte et micro-organisme) montre une certaine homogénéité. Pourtant, la description détaillée de toutes les interactions spécifiques

entre un hôte et un micro-organisme met à jour une grande hétérogénéité des mécanismes. Chaque micro-organisme a construit au cours de son évolution sa propre stratégie de manipulation de l'hôte pour optimiser sa survie et sa transmission. Ainsi, quand un micro-organisme se retrouve face à un hôte naïf, il entraîne souvent une réponse inadaptée de la part de l'hôte [80].

Les puces à ADN offrent la possibilité de décrire de manière précise la réponse de l'hôte face à un micro-organisme. Elles examinent les composantes de la relation hôte-pathogène dans l'ensemble de leur dimension. Nos études ont exploré divers aspects de la réponse inflammatoire au niveau du transcriptome de l'hôte. La possible dichotomie entre inflammation d'origine infectieuse ou non sous-tend une grande part de nos recherches. Le même outil d'analyse a été utilisé sur des cellules en culture, dans des modèles animaux, ou *in-vivo* en pathologie humaine.

Jusqu'à récemment, le modèle de choix pour expliquer la morbidité et la mortalité liées au sepsis était celui d'une réponse hyperinflammatoire non contrôlée de la part de l'hôte [81]. Ce modèle était soutenu par de nombreuses études dans lesquelles les animaux recevaient des doses importantes de bactéries ou de produits bactériens. L'injection de quantités exagérées de bactérie entraîne la libération brusque d'un flot de médiateurs inflammatoires entraînant le décès de l'hôte [82-87]. De nombreux essais cliniques ont évalué le bénéfice de traitements anti-inflammatoires sur la mortalité. Ces essais ont été des échecs. Ces échecs ont conduit à reconsidérer l'hypothèse selon laquelle seule la réponse pro-inflammatoire de l'hôte était responsable de la mortalité du sepsis. De plus, des modèles pré-cliniques ont été

élaborés mimant de façon plus conforme le sepsis humain [88].

La relation entre l'hôte et un micro-organisme semble basée sur une réponse pro-inflammatoire précoce suivie de façon rapprochée par une réponse anti-inflammatoire, conduisant à un état d'immunosuppression en cas d'évolution défavorable [5]. La décortication de la relation hôte-pathogène permet d'avancer dans la compréhension des mécanismes physiopathologiques du sepsis.

La première variable à intégrer dans la modulation de la réponse de l'hôte est le **temps**. La réponse de l'hôte au pathogène est un processus dynamique. L'intégration du temps est fondamentale pour interpréter la pertinence de la mesure d'un gène comme marqueur diagnostic ou comme cible thérapeutique potentielle. Dans l'analyse, le temps s'intègre comme « le temps qui s'écoule à partir du début de l'agression », mais également comme « le moment de la journée où survient l'agression ».

Nous avons décrit la réponse de l'hôte au cours du temps après injection d'acide oléique en intraveineux dans un modèle murin [89]. Un millier de gènes sont modulés dans le parenchyme pulmonaire après l'agression. Six points de mesures ont été réalisés (1h, 1h30, 3h, 4h, 18h et 24h). Trois profils d'expression (*cluster*) ont été individualisés : précoce (1h-1h30), intermédiaire (3h-4h) et tardif (18h-24h). Ces trois *clusters* de gènes ont été associés à des fonctions différentes. Les gènes modulés précocement sont impliqués dans la réponse pro-inflammatoire (cytokines pro-inflammatoires, récepteurs et protéines impliquées dans la transduction du signal).

Ceux modulés entre 3 et 4 h sont associés à l'expression de molécules d'adhésions, de composants du cytosquelette et à la migration transendothéliale des cellules du système immunitaire. Ceux activés tardivement sont impliqués dans la transcription, la synthèse des protéines, et le métabolisme, notamment celui du cholestérol.

Nos résultats sont en accord avec un modèle humain basé sur l'administration d'endotoxine à des sujets sains [67]. Ces modèles sont comparables puisque, dans les deux cas, l'agression n'était pas létale. L'expression de plusieurs cytokines et chemokines pro-inflammatoires (TNF, IL1A, IL1B, CXCL1 (GRO-α), CXCL2 (GRO-β), CCL2 (MCP1), CXCL8 (IL8), CXCL10) atteint un pic entre 2 et 4 h après l'injection d'endotoxine. Les facteurs de transcription connus pour contrôler l'expression de ces gènes (NFKB1, NFKB2, RELA, RELB) sont également à leur pic d'expression dans cette période. Quatre à 6 h après l'injection de LPS, on observe une poursuite de l'activation de la réponse inflammatoire (par certains facteurs de transcription comme STAT et CREB). De façon concomitante, plusieurs mécanismes participent au contrôle de cette réponse (SOCS3, IKBK, IL1RAP, IL1R2, TNFRSF1A). Cette période de temps charnière est critique pour le retour à un phénotype normal. Nos données histologiques montraient un retour à un phénotype pulmonaire normal après 4 h. Il faut noter que cette temporalité de la réponse semble être dose-dépendante [90].

La traduction clinique de cette approche a été faite dans le projet intitulé Sepsi*Chip*. L'objectif primaire était de déterminer une signature transcriptionnelle de la pneumonie chez des patients traumatisés. Les patients traumatisés avaient été

choisis pour avoir un fonds inflammatoire commun dans l'ensemble de notre cohorte, permettant ainsi d'identifier des marqueurs spécifiques d'infection en éliminant le bruit de fond (inflammation). Aucune signature transcriptionnelle spécifique n'a été mise en évidence chez les patients développant une pneumonie. Les différences entre les patients au moment de l'admission et au moment du début de la pneumonie étaient liées au facteur temps, puisque les mêmes variations étaient retrouvées chez des sujets contrôles. Les seules variations retrouvées dans cette étude étaient une diminution de la production de médiateurs pro-inflammatoires lors du diagnostic de pneumonie. Ces données, en accord avec une étude précédemment publiée [70], démontrent que le facteur temps est essentiel dans l'analyse du transcriptome. Cette composante doit donc être prise en compte pour toute analyse de données de transcriptome.

Après infection par *Coxiella burnetii,* nous avons montré que l'expression hépatique des gènes impliqués dans le contrôle de l'horloge biologique (ARNTL, CLOCK, PER2) au niveau moléculaire est modifiée [91]. L'horloge circadienne est perturbée précocement, puis l'expression des gènes revient à un niveau d'expression basal. Parallèlement, on observe un pic d'expression d'IFN-γ et d'IL-6 à J4, alors que l'expression d'IL10 et de son effecteur (STAT3) augmente constamment durant les sept jours de l'étude. Les modulations du rythme circadien ne sont retrouvées que chez les souris femelles. Ceci souligne l'interaction des différentes variables (temps et sexe) dans la modulation de la réponse de l'hôte.

La littérature montre que le rythme circadien interfère avec la réponse de l'hôte.

La variation circadienne de l'expression des gènes qui composent l'horloge biologique au niveau moléculaire (comme CLOCK ou ARNTL) influe sur l'expression de cytokines [92] ou l'activité des cellules NK [93]. La réponse à une stimulation antigénique est différente selon le moment de la journée [94]. Après injection d'oxazolone (allergène utilisé pour tester l'hypersensibilité de type retardée) en sous-cutanée, le nombre de cellules immunitaires et l'inflammation (mesurée par l'épaisseur cutanée) diffèrent selon que l'injection est réalisée le jour ou la nuit. Une inversion du cycle jour/nuit alterne le profil de réponse par rapport à une heure fixe d'injection (à 10 h ou 22 h). De même, la susceptibilité de drosophiles infectées par *Pseudomonas aeruginosa* est maximale quand l'infection survient en milieu de nuit [95]. Ces données confirment l'interaction entre le rythme circadien et la réponse de l'hôte.

L''influence du **sexe** dans la réponse de l'hôte a été largement décrite mais reste difficile à apprécier. Actuellement, la prise en charge des patients en sepsis grave ne diffère pas en fonction du sexe. Dans d'autres champs d'application, comme la traumatologie crânienne ou l'arrêt cardiaque, des essais cliniques en cours évaluent l'effet de traitements par de la progestérone sur le pronostic des patients (Clinical Trials NCT00822900, NCT01143064, NCT00048646). Les femmes ont généralement une production accrue d'anticorps. Elles développent d'ailleurs plus de maladies de système que les hommes [96]. Par rapport aux hommes, elles ont une moindre susceptibilité au sepsis [76,97], nécessitent moins souvent d'être hospitalisées en réanimation [98] et développent moins de complications [98,99]. Par contre, les

données épidémiologiques montrent que la mortalité est équivalente entre les deux sexes. Ces différences sont supportées par des données *in-vitro* où des cellules immunitaires exposées *ex-vivo* à du LPS produisent plus de cytokines lorsqu'elles proviennent d'un homme que d'une femme [100,101]. Ceci explique la réaction inflammatoire plus soutenue chez les hommes.

Des données épidémiologiques ont montré que le sexe ratio des patients développant une fièvre Q est d'environ deux hommes pour une femme [102]. Le rôle des hormones sexuelles a été confirmé dans un modèle murin d'infection par *C. burnetii*. On dénombrait moins de bactéries et de granulomes dans le foie et la rate des souris femelles que chez les souris mâles [103]. Pour expliquer avec plus de précisions ces observations, nous avons élaboré une approche transcriptionnelle. En fait, 86 % des 2 777 gènes modulés par l'infection sont sous l'influence du sexe. Brièvement, chez les mâles, les gènes étaient rattachés à l'adhésion cellulaire, la production de défensines, la transduction du signal et l'activation de la voie Jak/Stat. Ceux modulés chez les femelles étaient associés au rythme circadien. Pour affiner le modèle, nous avons soumis les souris à une castration chirurgicale. Cela a entraîné une réduction de 60 % des gènes modulés. Ce résultat confirme donc un dimorphisme sexuel au niveau de la réponse transcriptionnelle et le rôle prépondérant joué par les hormones sexuelles. Ces résultats posent également les bases pour expliquer le dimorphisme sexuel observé dans les données épidémiologiques.

Dans l'analyse de la réponse à l'infection par le virus *Influenza* [104], cinq souches virales ont été utilisées. Ces souches variaient par leur caractère antigénique

(hémaglutinine et neuraminidase) et leur **virulence** [105]. Des modèles *in vitro [106]* ou *in vivo [107,108]* ont mis en évidence une forte augmentation de l'expression des gènes de réponse à l'IFN après infection par la souche H5N1. Cette induction est probablement liée à la virulence de cette souche [107,108]. Dans notre étude, l'infection par H5N1 entraîne une expression supérieure en nombre et en valeur absolue des gènes de la réponse inflammatoire par rapport à l'infection par les autres souches. Pour les gènes de réponse à l'IFN (OAS1, OASL, IFITM1, IRF7, ICAM1, G1P2, et IFIT3), on observait un ratio d'expression de 2 à 4 fois supérieur pour l'infection par H5N1, par rapport aux autres souches. Au contraire, l'infection par la souche A/New Caledonia/20/99 (H1N1) entraîne des modifications transcriptionnelles modestes. Ainsi, sur plus de 8 500 gènes testés, l'infection par H1N1 modulait 36 gènes, alors que les autres souches modulaient de 1 455 (H5N1) à 3 020 gènes (H7N1). Ces modifications obtenues avec H1N1 étaient corrélées à la faible efficacité de réplication de cette souche dans les cellules.

Le virus de la grippe est un exemple caricatural de la variabilité des facteurs impliqués dans la relation hôte-pathogène. Toutefois, bien que des déterminants intrinsèques contribuent fortement à la virulence d'un agent infectieux [109-111], la naïveté de l'hôte face à cet agent joue également un rôle dans la gravité du tableau clinique [112]. La complexité des déterminants génétiques de la virulence des micro-organismes a été bien décrite ailleurs [35]. Le rôle des facteurs de virulence dans la variabilité de la réponse de l'hôte est également rapporté dans l'infection bactérienne. Les souches pathogènes d'une espèce bactérienne diffèrent des souches commensales

par l'acquisition et l'expression de gènes de virulence particuliers [113]. Les toxines bactériennes participent pleinement aux lésions infligées à l'hôte. Le tableau le plus connu est celui du *toxic-shock-syndrome* lié aux infections à staphylocoque doré et streptocoque [114,115]. Ces toxines sont fréquemment codées par des gènes localisés dans des îlots génomiques de structure particulière et regroupant des gènes de virulence. Certains de ces îlots génomiques sont communs à de nombreuses bactéries [116]. La présence de ces îlots distingue souvent les souches pathogènes des souches non virulentes [117].

L'ensemble des travaux présentés dans cette thèse utilise la même technologie d'analyse du transcriptome par puces à ADN. Bien que les résultats d'expression des principaux gènes modulés aient été confirmés par une technologie alternative, cette approche présente plusieurs limitations importantes. Tout d'abord, les types de puces à ADN employés dans ces travaux ne permettent pas d'explorer les variants alternatifs exprimés à partir d'un même gène, variable importante de la régulation de l'expression des gènes. La rétro-transcription à l'aide d'amorces poly-dT élimine également de l'analyse les modulations possibles des ARN non-codants. D'autre part, la corrélation entre la modulation de la transcription et de la traduction n'est pas toujours évidente. En plus d'être traduite, de nombreuses protéines dépendent pour être fonctionnelles de modifications post-traductionnelles spécifiques et de leur transport dans des compartiments sub-cellulaires particuliers. Ces différentes étapes de la régulation de l'expression des gènes sont complètement ignorées par l'analyse présentée ici. Ces limitations sont importantes à considérer lorsque des hypothèses

physiopathologiques sont élaborées à partir d'une simple analyse du transcriptome. Le développement de technologies comme le RNA-seq (pour étudier les variants alternatifs des ARNm), l'intégration de l'analyse de l'expression des ARN non codants et le développement des technologies d'analyses du protéome permettent de compléter théoriquement ces limitations.

Conclusion et perspectives

Le transcriptome est un outil de choix pour analyser la réponse de l'hôte. Nous avons montré dans ces différentes études l'influence de facteurs liés à l'environnement (le temps), à l'hôte (le sexe) et au micro-organisme (la virulence) dans la relation hôte-pathogène. La prise en compte du temps est capitale dans l'interprétation des données transcriptionnelles. La réponse de l'hôte est variable en fonction du délai écoulé entre le début de l'infection et celui de la prise en charge. L'intérêt d'une antibiothérapie très précoce dans le sepsis grave s'inscrit probablement dans cette approche. De plus, la réponse du système immunitaire est modulée par le rythme circadien. Ceci ouvre des perspectives de recherche intéressantes, notamment en réanimation où les agents de la sédation, l'absence d'alternance lumière / obscurité et l'activité constante perturbent l'horloge biologique. Le premier facteur de variation interindividuelle est probablement le sexe. Plus de 80 % des gènes modulés dans l'infection par *C. burnetii* sont dépendants du sexe. Ces données expérimentales éclairent les observations épidémiologiques. En perspective, nous élaborons une étude similaire en utilisant un modèle murin d'infection plurimicrobiennes intra-abdominale. Ce modèle ferait ainsi le lien avec notre pratique clinique. La prise en compte du sexe est probablement une des premières étapes d'une prise en charge individualisée. Par ailleurs, des approches basées sur des traitements hormonaux sont en cours de développement dans diverses pathologies et mériteraient d'être évaluées dans le sepsis. Enfin, la virulence du micro-organisme a un rôle propre. Nous avons démontré qu'il existe une variabilité

de réponse en fonction des souches du virus de la grippe. Les deux partenaires de la relation hôte-pathogène ont donc un rôle dans cette interaction. En utilisant des logiciels comme « *TranscriptomeBrowser*», les données transcriptionnelles provenant de différents modèles pourront être analysées afin de déterminer à large échelle l'impact des facteurs de virulence des micro-organismes sur la réponse de l'hôte.

Références bibliographiques

1. Angus DC, Wax RS: **Epidemiology of sepsis: an update**. *Crit. Care Med* 2001, **29**:S109-116.

2. Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R, Reinhart K, Angus DC, Brun-Buisson C, Beale R, et al.: **Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008**. *Crit. Care Med* 2008, **36**:296-327.

3. Hotchkiss RS, Karl IE: **The pathophysiology and treatment of sepsis**. *N. Engl. J. Med* 2003, **348**:138-150.

4. Hotchkiss RS, Nicholson DW: **Apoptosis and caspases regulate death and inflammation in sepsis**. *Nat. Rev. Immunol* 2006, **6**:813-822.

5. Hotchkiss RS, Coopersmith CM, McDunn JE, Ferguson TA: **The sepsis seesaw: tilting toward immunosuppression**. *Nat. Med* 2009, **15**:496-497.

6. Fiedler VB, Loof I, Sander E, Voehringer V, Galanos C, Fournel MA: **Monoclonal antibody to tumor necrosis factor--alpha prevents lethal endotoxin sepsis in adult rhesus monkeys**. *J. Lab. Clin. Med* 1992, **120**:574-588.

7. Alexander HR, Sheppard BC, Jensen JC, Langstein HN, Buresh CM, Venzon D, Walker EC, Fraker DL, Stovroff MC, Norton JA: **Treatment with recombinant human tumor necrosis factor-alpha protects rats against the lethality, hypotension, and hypothermia of gram-negative sepsis**. *J. Clin. Invest* 1991, **88**:34-39.

8. Romulo RL, Palardy JE, Opal SM: **Efficacy of anti-endotoxin monoclonal antibody E5 alone or in combination with ciprofloxacin in neutropenic rats with Pseudomonas sepsis**. *J. Infect. Dis* 1993, **167**:126-130.

9. Osuchowski MF, Welch K, Siddiqui J, Remick DG: **Circulating cytokine/inhibitor profiles reshape the understanding of the SIRS/CARS continuum in sepsis and predict mortality**. *J. Immunol* 2006, **177**:1967-1974.

10. Osuchowski MF, Welch K, Yang H, Siddiqui J, Remick DG: **Chronic sepsis mortality characterized by an individualized inflammatory response**. *J. Immunol* 2007, **179**:623-630.

11. Gogos CA, Drosou E, Bassaris HP, Skoutelis A: **Pro- versus anti-inflammatory cytokine profile in patients with severe sepsis: a marker for prognosis and future therapeutic options**. *J. Infect. Dis* 2000, **181**:176-180.

12. Le Gall JR, Lemeshow S, Saulnier F: **A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study**. *JAMA* 1993, **270**:2957-2963.

13. Knaus WA, Draper EA, Wagner DP, Zimmerman JE: **APACHE II: a severity of disease classification system**. *Crit. Care Med* 1985, **13**:818-829.

14. Christie JD: **Microarrays**. *Crit. Care Med* 2005, **33**:S449-452.

15. Carninci P, Yasuda J, Hayashizaki Y: **Multifaceted mammalian transcriptome**. *Curr. Opin. Cell Biol* 2008, **20**:274-280.

16. Jacquier A: **The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs**. *Nat. Rev. Genet* 2009, **10**:833-844.

17. Klickstein LB: **Production of a subtracted cDNA library**. *Curr Protoc Mol Biol* 2001, **Chapter 25**:Unit 25B.1.

18. Blackshaw S, St Croix B, Polyak K, Kim JB, Cai L: **Serial analysis of gene expression (SAGE): experimental method and data analysis**. *Curr Protoc Hum Genet* 2007, **Chapter 11**:Unit 11.7.

19. Staal FJT, van der Burg M, Wessels LFA, Barendregt BH, Baert MRM, van den Burg CMM, van Huffel C, Langerak AW, van der Velden VHJ, Reinders MJT, et al.: **DNA microarrays for comparison of gene expression profiles between diagnosis and relapse in precursor-B acute lymphoblastic leukemia: choice of technique and purification influence the identification of potential diagnostic markers**. *Leukemia* 2003, **17**:1324-1332.

20. Katagiri F, Glazebrook J: **Overview of mRNA expression profiling using DNA microarrays**. *Curr Protoc Mol Biol* 2009, **Chapter 22**:Unit 22.4.

21. Nagalakshmi U, Waern K, Snyder M: **RNA-Seq: a method for comprehensive transcriptome analysis**. *Curr Protoc Mol Biol* 2010, **Chapter 4**:Unit 4.11.1-13.

22. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR: **Highly integrated single-base resolution maps of the epigenome in Arabidopsis**. *Cell* 2008, **133**:523-536.

23. Vivancos AP, Güell M, Dohm JC, Serrano L, Himmelbauer H: **Strand-specific deep sequencing of the transcriptome**. *Genome Res* 2010, **20**:989-999.

24. Cloonan N, Forrest ARR, Kolle G, Gardiner BBA, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, et al.: **Stem cell transcriptome profiling via massive-scale mRNA sequencing**. *Nat. Methods* 2008, **5**:613-619.

25. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW: **The antisense transcriptomes of human cells**. *Science* 2008, **322**:1855-1857.

26. Armour CD, Castle JC, Chen R, Babak T, Loerch P, Jackson S, Shah JK, Dey J, Rohl CA, Johnson JM, et al.: **Digital transcriptome profiling using selective hexamer priming for cDNA synthesis**. *Nat. Methods* 2009, **6**:647-649.

27. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A: **Transcriptome analysis by strand-specific sequencing of complementary DNA**. *Nucleic Acids Res* 2009, **37**:e123.

28. Croucher NJ, Fookes MC, Perkins TT, Turner DJ, Marguerat SB, Keane T, Quail MA, He M, Assefa S, Bähler J, et al.: **A simple method for directional transcriptome sequencing using Illumina technology**. *Nucleic Acids Res* 2009, **37**:e148.

29. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics**. *Nat. Rev. Genet* 2009, **10**:57-63.

30. Der SD, Zhou A, Williams BR, Silverman RH: **Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays**. *Proc. Natl. Acad. Sci. U.S.A* 1998, **95**:15623-15628.

31. Zhu H, Cong JP, Mamtora G, Gingeras T, Shenk T: **Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays**. *Proc. Natl. Acad. Sci. U.S.A* 1998, **95**:14470-14475.

32. Huang Q, Liu D, Majewski P, Schulte LC, Korn JM, Young RA, Lander ES, Hacohen N: **The plasticity of dendritic cell responses to pathogens and their components**. *Science* 2001, **294**:870-875.

33. Nau GJ, Richmond JFL, Schlesinger A, Jennings EG, Lander ES, Young RA: **Human macrophage activation programs induced by bacterial pathogens**. *Proc. Natl. Acad. Sci. U.S.A* 2002, **99**:1503-1508.

34. Feezor RJ, Oberholzer C, Baker HV, Novick D, Rubinstein M, Moldawer LL, Pribble J, Souza S, Dinarello CA, Ertel W, et al.: **Molecular characterization of the acute inflammatory response to infections with gram-negative versus gram-positive bacteria**. *Infect. Immun* 2003, **71**:5803-5813.

35. Jenner RG, Young RA: **Insights into host responses against pathogens from transcriptional profiling**. *Nat. Rev. Microbiol* 2005, **3**:281-294.

36. Chinnaiyan AM, Huber-Lang M, Kumar-Sinha C, Barrette TR, Shankar-Sinha S, Sarma VJ, Padgaonkar VA, Ward PA: **Molecular signatures of sepsis: multiorgan gene expression profiles of systemic inflammation**. *Am. J. Pathol* 2001, **159**:1199-1209.

37. Cobb JP, Laramie JM, Stormo GD, Morrissey JJ, Shannon WD, Qiu Y, Karl IE, Buchman TG, Hotchkiss RS: **Sepsis gene expression profiling: murine splenic compared with hepatic responses determined by using complementary DNA microarrays**. *Crit. Care Med* 2002, **30**:2711-2721.

38. Chung TP, Laramie JM, Meyer DJ, Downey T, Tam LHY, Ding H, Buchman TG, Karl I, Stormo GD, Hotchkiss RS, et al.: **Molecular diagnostics in sepsis: from bedside to bench**. *J. Am. Coll. Surg* 2006, **203**:585-598.

39. Liu Y, Lin J-dong, Xiao X-jian, Zhang B-lei, Lin H: **[An investigation of changes in gene expression profile of heart tissue in a rat sepsis model]**. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue* 2009, **21**:155-159.

40. Checchia PA, Schierding W, Polpitiya A, Dixon D, Macmillan S, Muenzer J, Stromberg P, Coopersmith CM, Buchman TG, Cobb JP: **Myocardial transcriptional profiles in a murine model of sepsis: evidence for the importance of age**. *Pediatr Crit Care Med* 2008, **9**:530-535.

41. dos Santos CC, Gattas DJ, Tsoporis JN, Smeding L, Kabir G, Masoom H, Akram A, Plotz F, Slutsky AS, Husain M, et al.: **Sepsis-induced myocardial depression is associated with transcriptional changes in energy metabolism and contractile related genes: a physiological and gene expression-based approach**. *Crit. Care Med* 2010, **38**:894-902.

42. Maresh JG, Xu H, Jiang N, Shohet RV: **In vivo transcriptional response of cardiac endothelium to lipopolysaccharide**. *Arterioscler. Thromb. Vasc. Biol* 2004, **24**:1836-1841.

43. Scott LK, Vachharajani V, Mynatt RL, Minagar A, Conrad SA: **Brain RNA expression in obese vs lean mice after LPS-induced systemic inflammation**. *Front. Biosci* 2004, **9**:2686-2696.

44. Drevets DA, Schawang JE, Dillon MJ, Lerner MR, Bronze MS, Brackett DJ: **Innate responses to systemic infection by intracellular bacteria trigger recruitment of Ly-6Chigh monocytes to the brain**. *J. Immunol* 2008, **181**:529-536.

45. Li Z-jun, Li Y-ping, Gai H-rong, Xue Y-lai, Feng X-zeng: **[Research of gene expression profile of liver tissue in rat sepsis model]**. *Zhongguo Wei Zhong Bing Ji Jiu Yi Xue* 2007, **19**:156-159.

46. Weighardt H, Mages J, Jusek G, Kaiser-Moore S, Lang R, Holzmann B: **Organ-specific role of MyD88 for gene regulation during polymicrobial peritonitis**. *Infect. Immun* 2006, **74**:3618-3632.

47. Yang IV, Alper S, Lackford B, Rutledge H, Warg LA, Burch LH, Schwartz DA: **Novel Regulators of the Systemic Response to Lipopolysaccharide (LPS)**

[Internet]. *Am. J. Respir. Cell Mol. Biol* 2010, doi:10.1165/rcmb.2010-0342OC.

48. Sander LE, Sackett SD, Dierssen U, Beraza N, Linke RP, Müller M, Blander JM, Tacke F, Trautwein C: **Hepatic acute-phase proteins control innate immune responses during infection by promoting myeloid-derived suppressor cell function**. *J. Exp. Med* 2010, **207**:1453-1464.

49. Yu S-L, Chen H-W, Yang P-C, Peck K, Tsai M-H, Chen JJW, Lin F-Y: **Differential gene expression in gram-negative and gram-positive sepsis**. *Am. J. Respir. Crit. Care Med* 2004, **169**:1135-1143.

50. Mei SHJ, Haitsma JJ, Dos Santos CC, Deng Y, Lai PFH, Slutsky AS, Liles WC, Stewart DJ: **Mesenchymal stem cells reduce inflammation while enhancing bacterial clearance and improving survival in sepsis**. *Am. J. Respir. Crit. Care Med* 2010, **182**:1047-1057.

51. McDunn JE, Turnbull IR, Polpitiya AD, Tong A, MacMillan SK, Osborne DF, Hotchkiss RS, Colonna M, Cobb JP: **Splenic CD4+ T cells have a distinct transcriptional response six hours after the onset of sepsis**. *J. Am. Coll. Surg* 2006, **203**:365-375.

52. Goldmann O, von Köckritz-Blickwede M, Höltje C, Chhatwal GS, Geffers R, Medina E: **Transcriptome analysis of murine macrophages in response to infection with Streptococcus pyogenes reveals an unusual activation program**. *Infect. Immun* 2007, **75**:4148-4157.

53. Bluth M, Lin Y-Y, Zhang H, Viterbo D, Zenilman M: **Use of gene expression profiles in cells of peripheral blood to identify new molecular markers of acute pancreatitis**. *Arch Surg* 2008, **143**:227-233; discussion 233-234.

54. Scott LK, Vachharajani V, Minagar A, Mynatt RL, Conrad SA: **Differential RNA expression of hepatic tissue in lean and obese mice after LPS-induced systemic inflammation**. *Front. Biosci* 2005, **10**:1828-1834.

55. Prucha M, Ruryk A, Boriss H, Möller E, Zazula R, Herold I, Claus RA, Reinhart KA, Deigner P, Russwurm S: **Expression profiling: toward an application in sepsis diagnostics**. *Shock* 2004, **22**:29-33.

56. Shanley TP, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Doctor A, Kalyanaraman M, Tofil NM, Penfil S, Monaco M, et al.: **Genome-level longitudinal expression of signaling pathways and gene networks in pediatric septic shock**. *Mol. Med* 2007, **13**:495-508.

57. Cvijanovich N, Shanley TP, Lin R, Allen GL, Thomas NJ, Checchia P, Anas N, Freishtat RJ, Monaco M, Odoms K, et al.: **Validating the genomic signature of pediatric septic shock**. *Physiol. Genomics* 2008, **34**:127-134.

58. Fredriksson K, Tjäder I, Keller P, Petrovic N, Ahlman B, Schéele C, Wernerman J, Timmons JA, Rooyackers O: **Dysregulation of mitochondrial dynamics and the muscle transcriptome in ICU patients suffering from sepsis induced multiple organ failure**. *PLoS ONE* 2008, **3**:e3686.

59. Wong HR, Cvijanovich N, Allen GL, Lin R, Anas N, Meyer K, Freishtat RJ, Monaco M, Odoms K, Sakthivel B, et al.: **Genomic expression profiling across the pediatric systemic inflammatory response syndrome, sepsis, and septic shock spectrum**. *Crit. Care Med* 2009, **37**:1558-1566.

60. Tang BMP, McLean AS, Dawes IW, Huang SJ, Lin RCY: **Gene-expression profiling of peripheral blood mononuclear cells in sepsis**. *Crit. Care Med* 2009, **37**:882-888.

61. Johnson SB, Lissauer M, Bochicchio GV, Moore R, Cross AS, Scalea TM: **Gene expression profiles differentiate between sterile SIRS and early sepsis**. *Ann. Surg* 2007, **245**:611-621.

62. Zahar J-R, Timsit J-F, Garrouste-Orgeas M, Français A, Vésin A, Descorps-Declere A, Dubois Y, Souweine B, Haouache H, Goldgran-Toledano D, et al.: **Outcomes in severe sepsis and patients with septic shock: Pathogen species and infection sites are not associated with mortality [Internet]**. *Crit Care Med* 2011, doi:10.1097/CCM.0b013e31821b827c.

63. Ramilo O, Allman W, Chung W, Mejias A, Ardura M, Glaser C, Wittkowski KM, Piqueras B, Banchereau J, Palucka AK, et al.: **Gene expression patterns in blood leukocytes discriminate patients with acute infections**. *Blood* 2007, **109**:2066-2077.

64. Tang BMP, McLean AS, Dawes IW, Huang SJ, Cowley MJ, Lin RCY: **Gene-expression profiling of gram-positive and gram-negative sepsis in critically ill patients**. *Crit. Care Med* 2008, **36**:1125-1128.

65. Talwar S, Munson PJ, Barb J, Fiuza C, Cintron AP, Logun C, Tropea M, Khan S, Reda D, Shelhamer JH, et al.: **Gene expression profiles of peripheral blood leukocytes after endotoxin challenge in humans**. *Physiol. Genomics* 2006, **25**:203-215.

66. Prabhakar U, Conway TM, Murdock P, Mooney JL, Clark S, Hedge P, Bond BC, Jazwinska EC, Barnes MR, Tobin F, et al.: **Correlation of protein and gene expression profiles of inflammatory proteins after endotoxin challenge in human subjects**. *DNA Cell Biol* 2005, **24**:410-431.

67. Calvano SE, Xiao W, Richards DR, Felciano RM, Baker HV, Cho RJ, Chen RO, Brownstein BH, Cobb JP, Tschoeke SK, et al.: **A network-based analysis of systemic inflammation in humans**. *Nature* 2005, **437**:1032-1037.

68. Payen D, Lukaszewicz A-C, Belikova I, Faivre V, Gelin C, Russwurm S, Launay J-M, Sevenet N: **Gene profiling in human blood leucocytes during recovery from septic shock**. *Intensive Care Med* 2008, **34**:1371-1376.

69. Lissauer ME, Johnson SB, Bochicchio GV, Feild CJ, Cross AS, Hasday JD, Whiteford CC, Nussbaumer WA, Towns M, Scalea TM: **Differential expression of toll-like receptor genes: sepsis compared with sterile inflammation 1 day before sepsis diagnosis**. *Shock* 2009, **31**:238-244.

70. McDunn JE, Husain KD, Polpitiya AD, Burykin A, Ruan J, Li Q, Schierding W, Lin N, Dixon D, Zhang W, et al.: **Plasticity of the systemic inflammatory response to acute infection during critical illness: development of the riboleukogram**. *PLoS ONE* 2008, **3**:e1564.

71. Cobb JP, Moore EE, Hayden DL, Minei JP, Cuschieri J, Yang J, Li Q, Lin N, Brownstein BH, Hennessy L, et al.: **Validation of the Riboleukogram to Detect Ventilator-Associated Pneumonia After Severe Injury [Internet]**. *Ann. Surg* 2009, doi:10.1097/SLA.0b013e3181b8fbd5.

72. Pachot A, Lepape A, Vey S, Bienvenu J, Mougin B, Monneret G: **Systemic transcriptional analysis in survivor and non-survivor septic shock patients: a preliminary study**. *Immunol. Lett* 2006, **106**:63-71.

73. Wong HR, Shanley TP, Sakthivel B, Cvijanovich N, Lin R, Allen GL, Thomas NJ, Doctor A, Kalyanaraman M, Tofil NM, et al.: **Genome-level expression profiles in pediatric septic shock indicate a role for altered zinc homeostasis in poor outcome**. *Physiol. Genomics* 2007, **30**:146-155.

74. Feezor RJ, Baker HV, Xiao W, Lee WA, Huber TS, Mindrinos M, Kim RA, Ruiz-Taylor L, Moldawer LL, Davis RW, et al.: **Genomic and proteomic determinants of outcome in patients undergoing thoracoabdominal aortic aneurysm repair**. *J. Immunol* 2004, **172**:7103-7109.

75. Angele MK, Frantz MC, Chaudry IH: **Gender and sex hormones influence the response to trauma and sepsis: potential therapeutic approaches**. *Clinics (Sao Paulo)* 2006, **61**:479-488.

76. Eachempati SR, Hydo L, Barie PS: **Gender-based differences in outcome in patients with sepsis**. *Arch Surg* 1999, **134**:1342-1347.

77. Tosi MF: **Innate immune responses to infection**. *J. Allergy Clin. Immunol* 2005, **116**:241-249; quiz 250.

78. Manger ID, Relman DA: **How the host "sees" pathogens: global gene expression responses to infection**. *Curr. Opin. Immunol* 2000, **12**:215-218.

79. van der Poll T, Opal SM: **Host-pathogen interactions in sepsis**. *Lancet Infect Dis* 2008, **8**:32-43.

80. Kellam P, Weiss RA: **Infectogenomics: insights from the host genome into infectious diseases**. *Cell* 2006, **124**:695-697.

81. Offenstadt G, Maury E, Guidet B: **[Physiopathology and new treatments of septic shock]**. *Presse Med* 1996, **25**:1459-1465.

82. Freudenberg MA, Galanos C: **Tumor necrosis factor alpha mediates lethal activity of killed gram-negative and gram-positive bacteria in D-galactosamine-treated mice**. *Infect. Immun* 1991, **59**:2110-2115.

83. Kiener PA, Marek F, Rodgers G, Lin PF, Warr G, Desiderio J: **Induction of tumor necrosis factor, IFN-gamma, and acute lethality in mice by toxic and non-toxic forms of lipid A**. *J. Immunol* 1988, **141**:870-874.

84. Waage A, Espevik T: **Interleukin 1 potentiates the lethal effect of tumor necrosis factor alpha/cachectin in mice**. *J. Exp. Med* 1988, **167**:1987-1992.

85. Rothstein JL, Schreiber H: **Synergy between tumor necrosis factor and bacterial products causes hemorrhagic necrosis and lethal shock in normal mice**. *Proc. Natl. Acad. Sci. U.S.A* 1988, **85**:607-611.

86. Tracey KJ, Lowry SF, Fahey TJ 3rd, Albert JD, Fong Y, Hesse D, Beutler B, Manogue KR, Calvano S, Wei H: **Cachectin/tumor necrosis factor induces lethal shock and stress hormone responses in the dog**. *Surg Gynecol Obstet* 1987, **164**:415-422.

87. Lehmann V, Freudenberg MA, Galanos C: **Lethal toxicity of lipopolysaccharide and tumor necrosis factor in normal and D-galactosamine-treated mice**. *J. Exp. Med* 1987, **165**:657-663.

88. Eichacker PQ, Parent C, Kalil A, Esposito C, Cui X, Banks SM, Gerstenberger EP, Fitz Y, Danner RL, Natanson C: **Risk and the efficacy of antiinflammatory agents: retrospective and confirmatory studies of sepsis**. *Am. J. Respir. Crit. Care Med* 2002, **166**:1197-1205.

89. Lesur I, Textoris J, Loriod B, Courbon C, Garcia S, Leone M, Nguyen C: **Gene expression profiles characterize inflammation stages in the acute lung injury in mice**. *PLoS ONE* 2010, **5**:e11485.

90. Vedder H, Schreiber W, Yassouridis A, Gudewill S, Galanos C, Pollmächer T: **Dose-dependence of bacterial lipopolysaccharide (LPS) effects on peak response and time course of the immune-endocrine host response in humans**. *Inflamm. Res* 1999, **48**:67-74.

91. Textoris J, Ban LH, Capo C, Raoult D, Leone M, Mege J-L: **Sex-related differences in gene expression following Coxiella burnetii infection in mice: potential role of circadian rhythm**. *PLoS ONE* 2010, **5**:e12190.

92. Arjona A, Sarkar DK: **The circadian gene mPer2 regulates the daily rhythm of IFN-gamma**. *J. Interferon Cytokine Res* 2006, **26**:645-649.

93. Arjona A, Sarkar DK: **Circadian oscillations of clock genes, cytolytic factors, and cytokines in rat NK cells**. *J. Immunol* 2005, **174**:7618-7624.

94. Pownall R, Kabler PA, Knapp MS: **The time of day of antigen encounter influences the magnitude of the immune response**. *Clin. Exp. Immunol* 1979, **36**:347-354.

95. Lee J-E, Edery I: **Circadian regulation in the ability of Drosophila to combat pathogenic infections**. *Curr. Biol* 2008, **18**:195-199.

96. McCombe PA, Greer JM, Mackay IR: **Sexual dimorphism in autoimmune disease**. *Curr. Mol. Med* 2009, **9**:1058-1079.

97. Schröder J, Kahlke V, Staubach KH, Zabel P, Stüber F: **Gender differences in human sepsis**. *Arch Surg* 1998, **133**:1200-1205.

98. Wichmann MW, Inthorn D, Andress HJ, Schildberg FW: **Incidence and mortality of severe sepsis in surgical intensive care patients: the influence of patient gender on disease process and outcome**. *Intensive Care Med* 2000, **26**:167-172.

99. Schröder J, Kahlke V, Book M, Stüber F: **Gender differences in sepsis: genetically determined?** *Shock* 2000, **14**:307-310; discussion 310-313.

100. Moxley G, Posthuma D, Carlson P, Estrada E, Han J, Benson LL, Neale MC: **Sexual dimorphism in innate immunity**. *Arthritis Rheum* 2002, **46**:250-258.

101. Moxley G, Stern AG, Carlson P, Estrada E, Han J, Benson LL: **Premenopausal sexual dimorphism in lipopolysaccharide-stimulated production and secretion of tumor necrosis factor**. *J. Rheumatol* 2004, **31**:686-694.

102. Tissot-Dupont H, Raoult D: **Q fever**. *Infect. Dis. Clin. North Am* 2008, **22**:505-514, ix.

103. Leone M, Honstettre A, Lepidi H, Capo C, Bayard F, Raoult D, Mege JL: **Effect of sex on Coxiella burnetii infection: protective role of 17beta-estradiol**. *J. Infect. Dis* 2004, **189**:339-345.

104. Josset L, Textoris J, Loriod B, Ferraris O, Moules V, Lina B, N'guyen C, Diaz J-J, Rosa-Calatrava M: **Gene expression signature-based screening identifies new broadly effective influenza a antivirals [Internet]**. *PLoS ONE* 2010, **5**.

105. Taubenberger JK, Morens DM: **1918 Influenza: the mother of all pandemics**. *Emerging Infect. Dis* 2006, **12**:15-22.

106. Schmolke M, Viemann D, Roth J, Ludwig S: **Essential impact of NF-kappaB signaling on the H5N1 influenza A virus-induced transcriptome**. *J. Immunol* 2009, **183**:5180-5189.

107. Cameron CM, Cameron MJ, Bermejo-Martin JF, Ran L, Xu L, Turner PV, Ran R, Danesh A, Fang Y, Chan P-KM, et al.: **Gene expression analysis of host innate immune responses during Lethal H5N1 infection in ferrets**. *J. Virol* 2008, **82**:11308-11317.

108. Baskin CR, Bielefeldt-Ohmann H, Tumpey TM, Sabourin PJ, Long JP, García-Sastre A, Tolnay A-E, Albrecht R, Pyles JA, Olson PH, et al.: **Early and sustained innate immune response defines pathology and death in nonhuman primates infected by highly pathogenic influenza virus**. *Proc. Natl. Acad. Sci. U.S.A* 2009, **106**:3455-3460.

109. Conenello GM, Zamarin D, Perrone LA, Tumpey T, Palese P: **A single mutation in the PB1-F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence**. *PLoS Pathog* 2007, **3**:1414-1421.

110. Seo SH, Hoffmann E, Webster RG: **Lethal H5N1 influenza viruses escape host anti-viral cytokine responses**. *Nat. Med* 2002, **8**:950-954.

111. Kawaoka Y, Webster RG: **Sequence requirements for cleavage activation of influenza virus hemagglutinin expressed in mammalian cells**. *Proc. Natl. Acad. Sci. U.S.A* 1988, **85**:324-328.

112. Glezen WP: **Emerging infections: pandemic influenza**. *Epidemiol Rev* 1996, **18**:64-76.

113. Novick RP: **Autoinduction and signal transduction in the regulation of staphylococcal virulence**. *Mol. Microbiol* 2003, **48**:1429-1449.

114. Proft T, Sriskandan S, Yang L, Fraser JD: **Superantigens and streptococcal toxic shock syndrome**. *Emerging Infect. Dis* 2003, **9**:1211-1218.

115. Ulrich RG: **Evolving superantigens of Staphylococcus aureus**. *FEMS Immunol. Med. Microbiol* 2000, **27**:1-7.

116. Hacker J, Kaper JB: **Pathogenicity islands and the evolution of microbes**. *Annu. Rev. Microbiol* 2000, **54**:641-679.

117. Houdouin V, Bonacorsi S, Bidet P, Bingen-Bidois M, Barraud D, Bingen E: **Phylogenetic background and carriage of pathogenicity island-like domains in relation to antibiotic resistance profiles among Escherichia coli urosepsis isolates**. *J. Antimicrob. Chemother* 2006, **58**:748-751.

Annexe 1

Compartmentalization of Inflammation in Mice Blood and Lung during Acute Lung Injury.

Isabelle Lesur, Séverine Garnier, Béatrice Loriod, **Julien Textoris**, Marc Leone, Pascal Rihet and Catherine Nguyen*

Manuscrit tel que soumis à la revue PlosONE.

Compartmentalization of Inflammation in Mice Blood and Lung during Acute Lung Injury.

Isabelle Lesur, Séverine Garnier, Béatrice Loriod, Julien Textoris, Marc Leone, Pascal Rihet and Catherine Nguyen*

Lesur Isabelle: UMRS928, TAGC, Inserm, Parc scientifique de Luminy, Marseille, France – Aix-Marseille University, Marseille, France

Garnier Séverine: UMRS928, TAGC, Inserm, Parc scientifique de Luminy, Marseille, France – Aix-Marseille University, Marseille, France

Loriod Béatrice: UMRS928, TAGC, Inserm, Parc scientifique de Luminy, Marseille, France – Aix-Marseille University, Marseille, France

Textoris Julien: UMRS928, TAGC, Inserm, Parc scientifique de Luminy, Marseille, France – Aix-Marseille University, Marseille, France – Service d'anesthésie et de réanimation, hôpital Nord, AP-HM, Université de la méditerranée, Marseille, France

Leone Marc: Service d'anesthésie et de réanimation, hôpital Nord, AP-HM, Université de la méditerranée, Marseille, France

Rihet Pascal: UMRS928, TAGC, Inserm, Parc scientifique de Luminy, Marseille, France – Aix-Marseille University, Marseille, France

Nguyen Catherine: UMRS928, TAGC, Inserm, Parc scientifique de Luminy, Marseille, France – Aix-Marseille University, Marseille, France

* Corresponding author: nguyen@tagc.univ-mrs.fr

1 – Abstract

In this work, we describe a global overview of critical events occurring in blood during lung inflammation using a transcriptional approach in a murine model. By comparing transcriptional data in lung and blood, we show that response to inflammation is compartmentalized in these organs. Female C57BL6/J mice received an intravenous oleic acid injection (OA) to induce an acute lung injury (ALI). Peripheral Blood Mononuclear Cell (PBMC) expression patterns were analyzed using a 9900 cDNA mouse microarray (MUSV29K). Our analysis revealed marked changes in pathways involving the immune and inflammatory responses and lipid metabolism. First, the pro-inflammatory phase is activated, showing a high expression level of genes involved in stress response, DNA damage, protein synthesis, inflammation and immune response. The second phase was related to the tissular spreading of inflammation. This phase was associated with high expression of genes involved in metabolism and transcription. At the last stage of inflammation, lipid metabolism was disturbed and DNA repair was reactivated. We compared the transcriptional responses to inflammation in peripheral blood and lung. Several pathways including immune response, inflammatory response and disruption of lipid metabolism were simultaneously activated in PBMCs and lung. Other pathways, such as DNA repair and autophagy, were specifically activated in PBMCs. Finally, transcription was activated earlier in PBMCs (3h to 4h after OA injection) than in lungs (18h to 24h after OA injection). These findings, taken together with our previous transcriptional study in lungs, show that the inflammatory response is compartmentalized.

2 – Introduction

Acute lung injury (ALI) and its most severe manifestation, the acute respiratory distress syndrome (ARDS), is a clinical syndrome defined by an acute hypoxemic respiratory failure accompanied by bilateral pulmonary infiltrates consistent with edema and normal cardiac filling pressures [1-2]. ALI has many etiologies, suggesting that numerous genetic, environmental, and developmental factors are involved in its progression. Human studies have provided important and descriptive

information about the evolution of the physiological and inflammatory changes in the lungs during ALI. This information has led to hypotheses about mechanisms of injury [1-2]. However, these hypotheses have been difficult to test in humans because of too many clinical variables and confounding factors.

Animal models provide a bridge between patients and the laboratory bench [3]. Indeed, despite some limitations, studies using animal models of ALI/ARDS remain essential because, to date, there is no substitute for animal models as a tool to generate information about the pathophysiology of lung injury and to test novel therapeutic interventions in complex biological systems. The oleic acid model (OA) was successfully used in mouse to reproduce ALI and set up a model of lung inflammation [4]. The OA model produces the basic characteristics of ALI: early and rapidly reversible patchy inflammatory lung injury with permeability changes and impairment in gas exchange and lung mechanics [3]. Oleic acid has a direct toxicity to endothelial cells. Endothelial injury is followed by epithelial injury. Injection of OA leads to membrane damage and necrosis [3]. A major advantage of this model is its reproducibility; the administration of the same dose of OA by the same route to different animals is followed by a reproducible pulmonary injury [3].

Following OA injection, the pathophysiological events differ from organ to organ, and between organs and peripheral blood. This leads to compartmentalization, which can be seen as a way for the entire organism to limit the inflammation to its initial site. When the organism is unable to prevent the spreading of inflammation, the inflammation becomes systemic. The cellular compositions of each compartment, the micro-environment, and the leukocyte recruitment rate have an important influence on local inflammation and response [5]. During inflammation, a subtle balance exists between pro- and anti-inflammatory mediators both locally and at the systemic level. This time-dependent balance is the consequence of a complex network of up- and down-regulating signals. Specific surrounding cells that differ from compartment-to-compartment play a key role in this modulation. Similarly, distance between the initial site of inflammation and distal anatomic compartments influences the timing

and amplitude of the inflammatory response [5-8]. We can distinguish the local site of inflammation, the systemic circulation and distant compartments. The most striking differences are between tissues and the blood compartment [5]. Blood components have a uniform distribution, wherever they are found within the organism. In contrast, in the lungs, each cellular type has a specific localization. This means that cells cannot be simultaneously at the same inflammatory stage throughout the organ.

Similarly, at the organism level, the same compartmentalization process leads to an overwhelming inflammation across the whole organism. The inflammation process is associated with a cascade of events. During de-compartmentalization, which correspond to the spreading of inflammation from the initial site of inflammation to other compartments, high levels of pro-inflammatory mediators produced locally and released into the bloodstream initiate remote organ injury as a consequence of organ cross-talk. Injured organs are a site of synthesis of inflammatory mediators and the associated increase in vascular and epithelial permeability favors the leaking of mediators from one compartment to another [9, 10].

While the high concentrations of anti-inflammatory mediators present in the bloodstream prevent the ignition of new inflammatory foci, their presence within tissues may not always be sufficient to prevent the initiation of a deleterious inflammatory response in the different compartments. For example, in multiple organ failure patients with ARDS, Douzinas *et al.* reported higher IL-1β levels in pulmonary capillary blood than in peripheral vein blood [11]. These results strongly suggest the production of this cytokine within the tissue and their release in the downstream vein.

It has been previously shown that the gene expression patterns were significantly different when studied simultaneously in two different organs during inflammation. A previous study compared Peripheral Blood Mononuclear Cells (PBMC) and spleen immune cell functions in a murine burn model [12]. This study shows the need to consider gene expression in blood and organs simultaneously. Indeed, the inflammation status observed at the PBMC level alone in organisms undergoing lung inflammation does not allow for correct assessment of the inflammatory status of

each organs of the whole organism. The actual immune status at the tissue level has to be taken into consideration.

We have previously characterized changes in gene expression over time following OA-induced lung inflammation [4]. Gene expression patterns in lung were analyzed using a 9900 cDNA mouse microarray (MUSV29K) and a set of 1000 genes whose expression level significantly changed during a 24h time-course was identified. Subsequent functional annotation allowed us to find pathways related to pro- and anti-inflammatory stages.

In the present study, our first objective was to use the OA-induced ALI model previously described [4] to characterize lung and blood inflammation, using a transcriptional approach. In doing so, we provided an important database for future transcriptional studies of diseases associating inflammation. Since transcriptional studies about sepsis do not discriminate between the expression profiles due to infection associated with inflammation from the expression profile only due to inflammation, next experiments should specify the differences between inflammation due to an infectious process and that due to a non-infectious process. The second objective was to identify which pathways are commonly activated or repressed in blood compared with lung and to compare the timing of this activation/repression. Third, we show that the response to lung inflammation is compartmentalized in these two organs.

3- Results

a- Response Outcome of the mice

After intravenous OA administration, we noticed a transitional prostration of the mice, occurring 1h30 after injection and lasting for several hours. The mice came to a standstill in one corner of the cage and started shivering. The survival rate was 100% after 24h incubation.

b- Yield of PBMC and RNA isolation

For each mouse, we collected 200 to 900 µL of blood. Isolation of PBMC using Ficoll 1.084 allowed us to recover an average quantity of $1.9x10^6$ PBMC for 600 µl of blood. On average, we extracted 1.92 µg of total RNA for each sample (figure S1). RNA amplification was necessary since at least 3 µg of total RNA are required for microarray hybridization. On average, we amplified each of the RNA samples 15 times resulting in 28 µg of amplified RNA (aRNA).

c- Microarray analysis identified a set of genes significantly differentially expressed in blood during lung inflammation

Since we used the same protocol as previously set up in the OA-induced ALI model described in [4], we can confirm that we induced ALI in mice considered in this study. In order to study transcriptional changes associated with the lung inflammatory process in blood, we searched for a set of genes differentially expressed in this tissue at different incubation times. To this purpose, we used the multi-class Significant Analysis of Microarrays (SAM) procedure [13], applying a false discovery rate of 5% (figure 1). The analysis yielded a set of 1,534 significantly differentially expressed genes (table S2). We performed unsupervised hierarchical clustering to group these genes based on the similarities of their expression profiles over the time-course (figure 2). The Cluster software [14] classified the 1,534 genes previously identified based on their expression profile and we organized them into 4 clusters (figure 2).

Genes extracted from cluster A (409 genes) were transiently down-regulated 3h to 4h after OA injection. Their expression level was high at 1h-1h30 and 18h–24h time-points. Relative to their expression levels at the earlier time-points (1h-1h30), genes from cluster B (257 genes) showed increased expression between 3h and 4h and at 18h–24h. This timing may correspond to the transitional phase between the pro- and anti-inflammatory responses to OA injection that we have previously demonstrated for this model [4]. At the earliest time-points of the experiment (1h-1h30), genes from cluster C (314 genes) were over-expressed. After the early inflammatory phase, genes from cluster D (554 genes) were highly over-expressed between 3h and 24h (figure 2).

d- Identification of time-dependent pathways in blood during lung inflammation

Gene Ontology (GO) terms related to DNA damage such as « Response to stress » ($p=1.8x10^{-2}$), « Response to DNA damage stimulus » ($p=3.9x10^{-2}$) and « DNA repair » ($p=3.4x10^{-2}$) were overrepresented in cluster A (table 1). These terms were grouped together by the Functional Annotation Clustering tool from the Database for Annotation, Visualization and Integrated Discovery (DAVID) in a single annotation cluster with an enrichment score of 1.45. In cluster A, we also found genes involved in aminoacid metabolism. Indeed, GO terms such as « Nitrogen compound metabolic process » ($p=1.7x10^{-2}$), « Amine metabolic process » ($p=3x10^{-2}$) and «Aminoacid metabolic process » ($p=5.2x10^{-2}$) clustered together with an enrichment score of 1.14 (table 1).

“Regulation of metabolism” and “Transcription” terms were strongly represented in cluster B (table 2). GO terms such as «Regulation of cellular metabolic process» ($p=7.5x10^{-4}$), « RNA metabolic process » ($p= 7.1x10^{-3}$) and «Biopolymer metabolic process» ($p=2.2x10^{-3}$) clustered together with genes involved in « Transcription » ($p=5.7x10^{-3}$), « Regulation of transcription » ($p=6.6x10^{-3}$) and « DNA-dependent transcription » ($p=8.6x10^{-3}$) in the first functional annotation cluster identified by the Functional Annotation Clustering tool from DAVID (enrichment score: 1.9) (table 2). We found, in the fourth annotation cluster (enrichment score: 1.09), several pathways involved in the spreading of inflammation such as the « Il17 signaling pathway » Biocarta pathway ($p=1.7x10^{-2}$), the « Cell adhesion molecules (CAM) » Kegg pathway ($p=9.9x10^{-2}$) and GO terms such as « Biological adhesion » ($p=9.8x10^{-2}$) and « External side of plasma membrane » ($p=1.8x10^{-2}$) (table 2).

Genes grouped in cluster C are involved in pathways related to « Immune responses » ($p=1.5x10^{-2}$) such as « Chemokine activity » ($p=2x10^{-2}$) and « Cytokine-cytokine receptor interaction » ($p=3.1x10^{-2}$) as well as pathways related to the « Inflammatory response » ($p=1.7x10^{-2}$). Once again, the Functional Annotation Clustering tool from DAVID clustered these terms in a single group with a high

enrichment score (1.63) (table 3). GO terms related to « Autophagy » ($p=1.6x10^{-2}$) clustered together in the third annotation cluster with a 1.55 enrichment score (table 3). As expected in inflammation response, immune and pro-inflammatory responses were quickly activated. These findings validate our previously published experimental model of inflammation [4].

Functional annotation analysis of genes belonging to cluster D showed an over-representation of GO terms related to lipid metabolism, such as « Cholesterol metabolic process » ($p=1.2x10^{-3}$), « Steroid metabolic process » ($p=1.1x10^{-2}$) and « Cellular lipid metabolic process » ($p=6.5x10^{-2}$). These pathways were clustered together by the Functional Annotation Clustering tool from DAVID in a single cluster with an enrichment score of 1.55 (table 4, figure 3).

Overall, our gene expression analysis revealed marked sequential changes in pathways: i) first, the pro-inflammatory phase is activated with high expression levels of genes involved in stress response, DNA damage, protein synthesis, inflammation and immune response; ii) then, inflammation is spreading into tissues marked by high expression of genes involved in metabolism and transcription; iii) at the last stage of inflammation, we observe that lipid metabolism is highly disturbed concomitant with the reactivation of DNA repair.

4- Discussion

This study made it possible to: i) identify a set of 1,534 significantly differentially expressed genes which correspond to the response to ALI in PBMC at the transcriptional level and ii) identify time-dependent functional pathways. By comparing these results with the response to ALI in the lung [4], we showed that the response to lung inflammation is compartmentalized in lung and blood.

a- Transcriptional response to ALI in PBMC

In blood, the injection of OA significantly altered the expression levels of 1,534 genes that are organized into four clusters based on their expression profiles over time.

The expression levels of genes involved in defensive processes such as inflammatory and immune responses was affected 1h to 1h30 after OA injection. Among them, the transcription of several chemokine genes such as *Cxcr3, Cxcl15, Ccr1, Ccl22, Ccl25, Ccl3, Ccl6* and *Ccl8* was rapidly activated. This mechanism probably causes the recruitment of specific leukocyte subpopulations to sites of tissue damage (table S3) [15]. We found significant activation of the common pro-inflammatory cytokine Il1 as well as members of the Il1 receptor family (Il1r1, Il1r2).

The transcription of genes involved in DNA damage and DNA repair is activated at early time-points (1h-1h30) and reactivated at late points (18h–24h) of our experiments. This bimodal activation may be linked to cytokines' activity such as Il1 and Tnf which, in pro- and anti- inflammatory conditions, has been shown to induce nitric oxide synthase via the endothelial cells [16]. The nitric oxide that is then produced has been postulated to cause DNA damage by direct oxidization, resulting in mutations [17]. Genes involved in the Il17 signaling pathway, such as *Cd2* and *Cd34*, activated 3h to 4h after OA injection [18], may also be responsible for the re-activation of genes involved in DNA repair. Indeed, fibroblasts and other cells stimulated by Il17 are themselves induced to produce inflammatory and hematopoietic cytokines. Several distinct DNA repair proteins are involved in selective repair pathways. Oxidative DNA damage is predominantly repaired by nucleotide excision repair proteins such as Ercc3 and Rad23b. Other repair pathways, however, are also activated: the expression levels of *Neil3, Mre11a, Wrnip1* and *Mus81* increased; *Fanca, Ube2a* and *Ube2b*, which belong to the post-replication repair pathway, were highly expressed; and, similarly, transcription levels of genes, such as Msh3, involved in the DNA mismatch repair pathway was activated at early and late time-points.

We also found several genes involved in autophagy, such as Map1lc3a, Wipi1, Atg5 and Atg4b, highly expressed soon after OA injection (1h-1h30). Autophagy is an adaptive response to stress, a regulating process used by cells in order to recycle non-essential, redundant, or inefficient components. Simultaneously, it enables the stressed cell to gain vital nutrients. Autophagy is also involved in elimination of

intracellular microorganisms, tumor suppression, and antigen presentation [19].

Over all, at the early and later times points after injection, the physiological activity is focused on an adaptive response to the stress.

The transition phase, 3h to 4h after OA injection, is marked by genes involved in the spreading inflammation. Usually, the visualization of this process is not well-depicted because it occurs in a very short time between activation of the pro- and the anti-inflammatory immune responses. The relatively late activation of genes involved in the spreading of inflammation compared with genes involved in inflammation and immune processes is probably due to the fact that the vascular endothelium is the target of pro-inflammatory cytokines [20]. Once the vascular endothelium is activated, genes involved in cell adhesion are expressed and leukocytes are transferred from circulating blood to the site of inflammation. In our data, we observed the transitional activation of genes involved in leukocyte transendothelial migration and cell adhesion molecules (CAM), such as Cd34, Cd2, Cx3cl1 and Pecam1 (table S3). These molecules have been previously shown to be strongly involved in the inflammatory process since they modulate leukocyte trafficking allowing the circulation of inflammatory signals between organs [21-23].

The transitional phase is also associated with the activation of genes involved in primary metabolism and transcription. This is due to the cytokines, which were highly expressed at the early time-points, and which activated these pathways.

Finally, we found a late disruption of gene expression level related to the lipid metabolism. We identified 26 genes annotated to “lipid metabolic process” whose expression level increased from 3h to 4h after OA injection and reached a maximum level after 18h to 24h OA incubation (figure 3). Previous studies showed the disturbance of lipid metabolism soon after the initiation of inflammation or associated with inflammation and infection [4, 24-25]. Most of the changes in genes related to lipid metabolism that are induced by inflammation are attributable to the high transcription level of pro-inflammatory cytokines early after OA injection [24, 26-28] (table 1). Our findings on the timing of the effect of cytokines on lipid metabolism are in line with those obtained by Feingold *et al.* [29] .

We expected a massive release of inflammatory mediators soon after OA injection, followed by the release of anti-inflammatory mediators [28]. This sequence of events has been previously reported in models of sepsis [30].

Overall, at this complex stage (between 3h-4h) we observed elements links to the inflammation spreading, the activation of the primary metabolism and transcription, and lipid metabolism modulation, the lipid metabolism being mainly modulated at the latest time points. In this study, we confirmed the findings of our previous study [4]. We identified pathways activated simultaneously or delayed in either lungs or PBMC. Even in the same OA-induced ALI model in mouse, transcriptional profiles differ from one compartment to another at the same time-points. We show here that the response to inflammation is compartmentalized in mouse in a least two organs: lung and blood.

b- Compartmentalization of transcriptional response to ALI in lung and blood

We previously used the same mouse inflammation model to analyze the transcriptional profile in inflamed lung following OA injection [4]. However, lung and PBMC studies were not performed on the same mice. Given the robustness of this model, this limitation does not prevent to compare the response to lung inflammation in lung and PBMC (figure 4). We identified several pathways simultaneously activated in lung and PBMC, whereas other pathways are specifically activated in either one or the other organ. When the same pathways were activated in both lung and PBMC, their activation was not necessarily simultaneous. These differences imply that inflammation is compartmentalized in these two organs. Of the 1,534 genes whose expression changed significantly over time in PBMC, 720 (47%) are also significantly differentially expressed over time in the lung. The expression level variations of these 720 genes, which represent 72% of the 1,000 genes whose expression levels changed over the time-course in lung, are similar in both organs.

Among the similarities between the lung and PBMC responses, we found a simultaneous activation of the inflammation and immune responses. The expression levels of numerous chemokines increased early after OA injection, corresponding to

the pro-inflammatory response to ALI. Right after the pro-inflammatory phase, the activation of pathways involved in the spreading of inflammation such as the CAM proteins occurred simultaneously in PBMC and lung cells. This activation is concomitant with the high transcriptional rate of genes involved in the hematopoietic cell lineage pathway. Finally, we noted the simultaneous late disruption of genes, with a slight shift, related to lipid metabolism in both blood and the lung (18h to 24h after OA injection). Besides their role in lipid transport, lipoproteins seem to participate in innate immunity [4,24-25].

The shift in response to inflammation between lung and PBMC signed a de-compartmentalization process. Indeed, the PBMC response to lung inflammation shows several differences to that in the lung. First, more than 53% genes were significantly differentially expressed in PBMC over the time-course compared with lung (1,534 *vs*. 1,000). This suggests that the response to inflammation is more general in the PBMCs compared with lung. Out of the 1,534 genes, we identified 814 genes specifically differentially expressed in PBMC over time-course (*i.e.* 53%). In contrast, only 280 (28%) genes identified in the lung are differentially expressed only in this organ.

Among genes specifically overexpressed in PBMC, we identified genes involved in DNA damage and DNA repair at early (1h-1h30) and late time-points (18h-24h). Similarly, the activation of autophagy seems to be blood-specific. We suggest that the activations of the DNA repair and autophagy pathways are limited to the blood compartment because anti-inflammatory mediators already prevented the activation of these pathways at the lung level.

Finally, we showed that activation of transcription occurred earlier (3h to 4h after OA injection) in blood than in lung (18h to 24h after OA injection). This delay may be explained by the time required for the circulating inflammatory mediators to cross the lung border and by the fact that blood is the first organ in contact with OA.

Studies have demonstrated that under non-pathological conditions the microenvironment and tissue compartment (blood, liver, gut, spleen...) markedly influence immune cell functional capabilities, thereby creating diversity and

heterogeneity within the immune system [31-32]. Two previous studies compared the response of immune cells from various tissues after thermal injury and inflammation (spleen, blood, mesenteric lymph nodes, Peyer's patches) [12, 33]. They observed significant compartmental differences in the kinetics of immunosuppression. Likewise, the ability of PBMC to generate a pro-inflammatory response to endotoxin differs from that of splenic macrophages. Splenic macrophages from burned mice produced significantly higher levels of nitric oxide, PGE_2, TNFα and Il6 as compared with macrophages from control animals [12]. In contrast, PBMC isolated from injured mice only produced significantly elevated levels of PGE_2 as compared with those isolated from control mice [12]. Moreover, PBMC productive capacity for TNFα and Il6 was significantly suppressed at post-burn. This confirms that the timing of the pre-existing inflammation (pro- or anti-inflammatory stage) influences the outcome of the organism undergoing infection as well.

5- Conclusion

In this study, we assessed gene expression profiles in mouse PBMC associated with a lung inflammation during a 24h time-course. By comparing similarities and differences between the transcriptional profiles in lung and PBMC, we pointed out significant compartmental differences in the kinetics of the lung inflammation in mice as well as the consequences of de-compartmentalization for distant organs.

Overall, our microarray analysis provides a global overview of critical events occurring in blood within lung inflammation. We confirm that, in blood, gene-expression profile discriminates between pro- and anti-inflammatory phases during this inflammatory process. The analysis of gene functional annotation reveals several major features. As expected, it indicates that the pro-inflammatory stage is characterized by the activation of the immune response and inflammatory mechanisms. The ultimate severity of the inflammation depends on the interplay between pro- and anti-inflammatory mediators and repair mechanisms, beginning after the initiation of injury. Later, the immune cells migrate into tissues through interaction with vascular endothelial cells, metabolism regulation and transcription

are then activated. At the late stage of inflammation, metabolism is deeply disturbed. We showed here that inflammation is associated with marked changes in lipid and lipoprotein metabolism.

Our findings suggest that, because of the compartmentalization of the response to lung inflammation in blood and lung, changes of cell transcription profiles depend on the targeted organ. Thus, changes induced by lung inflammation in lung and PBMC compartments should be considered together. The timing of inflammation and the interaction between pro- and anti-inflammatory phases depend of each organ. The structure of each organ constitutes a barrier to the propagation of inflammation. It is responsible for the heterogeneity of the inflammatory response at the level of each organ. We show that the assessment of PBMC function alone in patients with lung inflammation may not accurately reflect their actual immune status. Nonetheless, future animal studies of lung inflammation should compare the analysis of PBMC and fixed tissue immune cell function at multiple time-points. Multi-parameter experimental analysis should improve the potential extrapolation of the experimental findings to the clinical arena.

6- Materials and Methods

a – Experimental groups

Wild-type female C57Bl/6J mice, 7 weeks old, were obtained from Janvier and housed in a specific pathogen-free animal facility. C57BL/6J mice have been chosen because it is the most widely used inbred strain. It has been used in a wide variety of research areas, especially in studies related to transcriptional response to pathogens. However, strain of mice may affect the transcriptomic response, as shown by Cavarra *et al.* [34]. We studied 37 mice divided into 6 groups (figure 5). Each group included one or two mice given physiological serum and 4 to 6 mice given OA. Each group was identified according to the incubation time: 1h, 1h30, 3h, 4h, 18h and 24h. During our previous work on lung expression profiles during inflammation, we performed real-time PCR on TNFα, Il10, Il4 and Il6 to identify the OA incubation time corresponding to the pro- and the anti-inflammatory stages [4]. We found that, in this model, the pro-inflammatory

stage occurred 1h to 1h30 after OA injection and the anti-inflammatory stage reached a maximum level 18h to 24h after OA injection. This work explains how we defined our time-points. All experimental procedures involving animals were approved by the veterinary office of the Ministry of Agriculture, France (authorization number: 13-27).

b- Injections

Female C57Bl/6J mice were anesthetized with 5% Xylazine – 20% Ketamine (0,1ml/10g). We injected 200 µl of physiological serum in the control mouse at each time-point and 200 µl of OA (1,2 µl/g body weight, sigma #27728-1L-R) in the other 26 mice of our study. Administration of physiological serum and OA was done through the tail vein with a 0.3-ml insulin syringe (BD #320837).

c- Blood sampling and PBMC isolation

Each group of mice was sacrificed 1h, 1h30, 3h, 4h, 18h or 24h after injection of OA or physiological serum by exsanguination through the eye vein. 200 to 900 µL of whole blood was collected in tubes containing 100µL EDTA 0.5M. PBMC were isolated using a ficoll gradient 1.084 (Ficoll-Paque Premium 1.084, GE Healthcare #17-5446-02). Cells were resuspended in 500 µL PBS and counted.

d- RNA extraction and amplification

Total RNA from PBMC was isolated using the Mouse RiboPure-Blood RNA isolation kit (Ambion, #AM1951). The quality of RNA was confirmed on the Agilent 2100 Bioanalyzer (Agilent, #G2938C) using the RNA 6000 Nano chips (Agilent, #5067-1511), and the concentration of RNA was determined by reading absorbance at 260/280 nm. RNA was amplified using the MessageAmp II aRNA amplification Kit (Ambion, #AM1751). For up to 5 µg RNA amplification, we used a 20 µl reverse transcription (RT) reaction with 1µl T7 Oligo(dT) Primer, 4µl dNTP, 1µl Ribonuclease inhibitor and 1µl Reverse transcriptase in 1× first-strand buffer with a 42°C incubation for 2 hours. Second-strand synthesis was carried out in 100 µl with 2µl DNA polymerase, 1µl RNase H, and 4µl dNTP in 1× second-strand buffer simply by adding 80 µl of an ice-cold second-strand premix to the ice-cold 20 µl RT

reaction and incubating at 16°C for 2 h. cDNA was purified on cDNA Filter Cartridge and was transcribed in 40 µl [37°C, 14 h, 4µl T7 RNA enzyme Mix, 7.5mM T7-GTP, 7.5mM T7-ATP, 7.5mM T7-UTP, 7.5mM T7-CTP, T7 1× buffer]. Amplified antisense RNA (aRNA) was purified on aDNA Filter Cartridge.

e- Microarray design

We used the GPL9478 Nylon platform described in the Gene Expression Omnibus (GEO) public repository [35-36]. This microarray contains 1Kb PCR fragments representing 7,771 bacterial clones matching 6622 mouse genes. 73% (4833) of the genes are represented by a single cDNA clone and about 27% (1789) of the genes included in this gene set are represented by two or more different cDNA clones, providing internal controls to assess the reproducibility of gene expression measurements. MUSV29K contains 8 positive controls (poly-A LBP2S and Cot1, a mix of DNA fragments containing repeated sequences) and 2,205 negative controls (empty spots and CG03, an *Arabidopsis thaliana* cDNA sequence). In summary, the MUSV29K mouse microarray contains 9,984 spots, identifying 6,622 mouse genes.

f- Microarray data acquisition and analysis

Twenty-five mRNA samples (one per mouse) were run on a single microarray. In addition, 12 samples were run on two microarrays, and were considered as technical replicates. This corresponds to a total of 49 microarrays (*i.e*. 37 mice). All microarray procedures were done at our microarray core facility [37]. cDNAs were designed and prepared as described in Puthier *et al.* [38], using 5µg of amplified mRNA in the presence of α-dCTP 33P.

After image acquisition, data were processed as described in figure 1. Hybridization signals were quantified using the Bzscan2 software [39]. All images were carefully inspected and spots with overestimated intensities due to neighborhood effects were manually excluded. We used the NylonArray library for R locally developed by A. Bergon and D. Puthier [40]. This package contains functions to perform diagnosis and normalization of nylon microarray data. For each microarray, we obtained two datasets: the first was obtained by hybridizing the

microarray with a probe, whose oligonucleotide sequence is common to all spotted PCR products (vector) and the second was obtained by hybridizing the microarray with PBMC cDNA (sample). The sample datasets were corrected for neighborhood effects and local background as described by F. Lopez *et al.* [39]. Quantile normalization was applied to sample data to correct for global intensity and dispersion.

Microarray data were statistically analyzed using the TIGR MeV (MultiExperiment Viewer) V4.3.01 software [41]. Significant Analysis of Microarrays (SAM) procedure was applied to look for time specific variation in gene expression in the dataset [13]. At each time-point, gene expressions in control samples (physiological serum injection) were subtracted from gene expressions in OA samples. Before partitioning the genes, we performed a hierarchical clustering of the samples in TMEV V4.3.01 using average linkage clustering metrics and Pearson correlation for the distance. This analysis partitioned the samples into three groups: group 1 (1h, 1h30), group 2 (3h, 4h) and group 3 (18h, 24h). Hierarchical clustering (average linkage clustering metrics and Pearson correlation for the distance) was then applied to the genes using the Cluster software and results were visualized with the Treeview software [14]. We identified biological annotations for the clusters using DAVID [42-43] and interaction networks using the Kegg pathways [44]. To interpret our data, we used two functionalities in DAVID: the « Functional annotation Chart» and the « Functional annotation clustering » tools. The first associates a gene ID with a biological term, which belongs to one of the 40 annotation categories available in DAVID, including GO terms, protein-protein interactions, protein functional domains, disease association and sequence features. This extended annotation coverage increases analytic power by allowing investigators to analyze their genes from many different biological aspects at once. DAVID functional annotation clustering measures relationships between annotation terms based on the number of genes sharing these terms. This type of grouping of functional annotations can provide greater insight into the relationships between annotation categories and terms than the traditional linear list of enriched terms, as it is not confounded by highly

related or semantically redundant terms. Each cluster of functional annotations was associated with an enrichment score, which depended on the distribution of the enrichment (p-values) of each member. A good enrichment score was obtained when most members of a cluster had significant p-values. This is a relative score and not a statistical probability, and has no minimum or maximum limits. This means that enrichment scores could be considered only together, by comparing them.

All data are contain the Minimum Information About a Microarray Experiment, meaning they are MIAME compliant and have been submitted to the GEO database [36] (accession number GSE19030).

7- Acknowledgements

We thank Valérie Gall for technical support and Aurélie Bergon for informatics support. This work was supported by grants from Conseil Régional Provence Alpes Côte d'Azur, Société Française d'Anesthésie et de Réanimation, and le Ministère de la santé (Programme Hospitalier de Recherche Clinique, PHRC 2005).

8- References

1. Ashbaugh DG, Bigelow DB, Petty TL, Levine BE (1967) Acute respiratory distress in adults. Lancet 2: 319-323.
2. Ware LB, Matthay MA (2000) The acute respiratory distress syndrome. N. Engl. J. Med 342: 1334-1349.
3. Matute-Bello G, Frevert CW, Martin TR (2008) Animal models of acute lung injury. Am. J. Physiol. Lung Cell Mol. Physiol 295: L379-399.
4. Lesur I, Textoris J, Loriod B, Courbon C, Garcia S, Leone M, Nguyen C (2010) Gene expression profiles characterize inflammation stages in the acute lung injury in mice. PLoS One 5(7):e11485.
5. Cavaillon JM, Annane D (2006) Compartmentalization of the inflammatory response in sepsis and SIRS. J Endotoxin Res 12(3):151-70.
6. Cavaillon J-M, Adib-Conquy M (2007) Compartmentalized activation of immune cells during sepsis and organ dysfunction IN "Update in Intensive Care and Emergency Medicine" –

"Mechanisms of sepsis induced organ dysfunction and recovery" (E. Abraham & M. Singer Eds) Springer-Verlag, Heidelberg; Volume 44, Part 3: 161-182.

7. Ipaktchi K, Mattar A, Niederbichler AD, Hoesel LM, Vollmannshauser S, Hemmila MR, Su GL, Remick DG, Wang SC, Arbabi S. (2006) Attenuating burn wound inflammatory signaling reduces systemic inflammation and acute lung injury. J Immunol. 177(11):8065-71.
8. Damazo AS, Yona S, Flower RJ, Perretti M, Oliani SM. (2006) Spatial and temporal profiles for anti-inflammatory gene expression in leukocytes during a resolving model of peritonitis. J Immunol. 176(7):4410-8.
9. Tutor JD, Mason CM, Dobard E, Beckerman RC, Summer WR, et al. (1994) Loss of compartmentalization of alveolar tumor necrosis factor after lung injury. Am. J. Respir. Crit. Care Med 149: 1107-1111.
10. Kurahashi K, Kajikawa O, Sawa T, Ohara M, Gropper MA, et al. (1999) Pathogenesis of septic shock in Pseudomonas aeruginosa pneumonia. J. Clin. Invest 104: 743-750.
11. Douzinas EE, Tsidemiadou PD, Pitaridis MT, Andrianakis I, Bobota-Chloraki A, et al. (1997) The regional production of cytokines and lactate in sepsis-related multiple organ failure. Am. J. Respir. Crit. Care Med 155: 53-59.
12. Schwacha MG, Schneider CP, Chaudry IH (2002) Differential expression and tissue compartmentalization of the inflammatory response following thermal injury. Cytokine 17: 266-274.
13. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U.S.A 98: 5116-5121.
14. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A 95: 14863-14868.
15. Rossi D, Zlotnik A (2000) The biology of chemokines and their receptors. Annu. Rev. Immunol 18: 217-242.
16. Geller DA, Nussler AK, Di Silvio M, Lowenstein CJ, Shapiro RA, Wang SC, Simmons RL, Billiar TR (1992) Cytokines, endotoxin and glucocorticoids regulate the expression of inducible nitric oxide synthase in hepatocytes. Proc.

Natl. Acad. Sci. USA, 90: 522–526.

17. Tamir S, Burney S, Tannenbaum SR (1996) DNA damage by nitric oxide. Chem. Res. Toxicol., 9: 821– 827.
18. Opdenakker G (2001) New insights in the regulation of leukocytosis and the role played by leukocytes in septic shock. Verh. K. Acad. Geneeskd. Belg 63: 531-538; discussion 538-541.
19. Nishida Y, Arakawa S, Fujitani K, Yamaguchi H, Mizuta T, et al. (2009) Discovery of Atg5/Atg7-independent alternative macroautophagy. Nature 461: 654-658.
20. Galley HF, Webster NR (2004) Physiology of the endothelium. Br J Anaesth. 93(1):105-13.
21. Muller WA (2009) Mechanisms of transendothelial migration of leukocytes. Circ. Res 105: 223-230.
22. Leone M, Garcin F, Chaabane W, Boutière-Albanèse B, Albanèse J, et al. (2003) Activation of adhesion molecules in patients with septic shock. Ann Fr Anesth Reanim 22: 721-729.
23. Woodfin A, Voisin M, Nourshargh S (2007) PECAM-1: a multi-functional molecule in inflammation and vascular biology. Arterioscler. Thromb. Vasc. Biol 27: 2514-2523.
24. Hardardóttir I, Grunfeld C, Feingold KR (1995) Effects of endotoxin on lipid metabolism. Biochem. Soc. Trans 23: 1013-1018.
25. Argilés JM, Lopez-Soriano FJ, Evans RD, Williamson DH (1989) Interleukin-1 and lipid metabolism in the rat. Biochem. J 259: 673-678.
26. Grunfeld C, Soued M, Adi S, Moser AH, Fiers W, et al. (1991) Interleukin 4 inhibits stimulation of hepatic lipogenesis by tumor necrosis factor, interleukin 1, and interleukin 6 but not by interferon-alpha. Cancer Res 51: 2803-2807.
27. Baumann H, Prowse KR, Marinković S, Won KA, Jahreis GP (1989) Stimulation of hepatic acute phase response by cytokines and glucocorticoids. Ann. N. Y. Acad. Sci 557: 280-295, discussion 295-296.
28. Hotchkiss RS, Coopersmith CM, McDunn JE, Ferguson TA (2009) The sepsis

seesaw: tilting toward immunosuppression. Nat. Med 15: 496-497.

29. Feingold KR, Staprans I, Memon RA, Moser AH, Shigenaga JK, et al. (1992) Endotoxin rapidly induces changes in lipid metabolism that produce hypertriglyceridemia: low doses stimulate hepatic triglyceride production while high doses inhibit clearance. J. Lipid Res 33: 1765-1776.
30. Johnson SB, Lissauer M, Bochicchio GV, Moore R, Cross AS, et al. (2007) Gene expression profiles differentiate between sterile SIRS and early sepsis. Ann. Surg 245: 611-621.
31. Takahashi K, Naito M, Takeya M (1996) Development and heterogeneity of macrophages and their related cells through their differentiation pathways. Pathol. Int 46: 473-485.
32. Murphy KM, Ouyang W, Farrar JD, Yang J, Ranganath S, et al. (2000) Signaling and transcription in T helper development. Annu. Rev. Immunol 18: 451-494.
33. Deitch EA, Xu DZ, Qi L (1990) Different lymphocyte compartments respond differently to mitogenic stimulation after thermal injury. Ann. Surg 211: 72-77.
34. Cavarra E, Fardin P, Fineschi S, Ricciardi A, De Cunto G, Sallustio F, Zorzetto M, Luisetti M, Pfeffer U, Lungarella G, Varesio L (2009) Early response of gene clusters is associated with mouse lung resistance or sensitivity to cigarette smoke. Am J Physiol Lung Cell Mol Physiol. 296(3):L418-29.
35. Barrett T, Edgar R (2006) Gene Expression Omnibus (GEO): Microarray data storage, submission, retrieval, and analysis. Methods Enzymol 411: 352-369.
36. GEO repository. Available at: http://www.ncbi.nlm.nih.gov/geo/
37. TAGC website: http://www.tagc.univ-mrs.fr/
38. Puthier D, Joly F, Irla M, Saade M, Victorero G, et al. (2004) A general survey of thymocyte differentiation by transcriptional analysis of knockout mouse models. J. Immunol 173: 6109-6118.
39. Lopez F, Rougemont J, Loriod B, Bourgeois A, Loï L, et al. (2004) Feature extraction and signal processing for nylon DNA microarrays. BMC Genomics 5: 38.
40. NylonArray: ftp://tagc.univ-mrs.fr/public.bzscan/nylonArray/

41. TMeV: http://www.tm4.org/mev.html
42. Sherman BT, Huang DW, Tan Q, Guo Y, Bour S, et al. (2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics 8: 426.
43. Lempicki RA, Sherman BT, Huang DW, Tan Q, Guo Y, et al. (2007) DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics 8: 426.
44. KEGG: http://www.genome.ad.jp/kegg/pathway.html

9- Figure Legends

Figure 1. Schematic outline of data analysis.

[a] We compared gene expression between three groups: 1h-1h30 / 3h-4h / 18h-24h.

[b] HCL: Hierarchical Clustering

We considered the full dataset (n=6,622 genes) to identify genes significantly differentially expressed between the three classes previously described. The 1,534 genes identified with the Significant Analysis of Microarrays (SAM) tool were clustered using the Cluster software and each cluster was independently functionally annotated using the Database for Annotation, Visualization and Integrated Discovery tool (DAVID).

Figure 2. Hierarchical classification of 1,534 significant genes differentially expressed during the time-course of inflammation.

This set of genes was extracted from the full dataset (n= 9,984) using a Significant Analysis of Microarrays (SAM) procedure and a false discovery rate of 5%. Each row represents a gene and each column represents a sample. Red and green indicate expression levels above and below the median respectively. Before partitioning the genes, we performed a hierarchical clustering of the samples in TMEV V4.3.01. Average linkage clustering metrics and Pearson correlation were

used for the distance. This analysis partitioned the samples into 3 groups based on the Oleic Acid (OA) incubation time (1h-1h30, 3h-4h, 18h-24h). Hierarchical clustering (average linkage clustering metrics and Pearson correlation for the distance) was then applied to the genes using the Cluster software and results were visualized with the Treeview software. 4 clusters were identified: A, B, C and D. Genes of cluster A (409 genes - blue) were transitively down-regulated 3h to 4h after OA injection. Their expression level was high at early time-points (1h-1h30) and late time-points (18h–24h). Genes of cluster B (257 genes - pink) were over-expressed between 3h and 4h relative to their expression level at the earlier time-points (1h-1h30) and the late measurements (18h–24h). At the earliest time-points of the experiment (1h-1h30), genes from cluster C (314 genes - green) were over-expressed. Following the early inflammatory phase, genes from cluster D (554 genes - yellow) were highly over-expressed between 3h and 24h. The dendrogram of genes, to the left of the matrix represents overall similarities in gene expression profiles.

Figure 3. Functional annotation of genes involved in lipid metabolism.

Of the genes in cluster D whose expression level increased 3h to 4h after Oleic Acid (OA) injection and reached a maximum level after 18h to 24h, 26 belong to the lipid metabolic process. These genes can be assigned to one or several of 4 Gene Ontology (GO) categories related to Lipid metabolism. “Grey” means the gene belongs to the pathway and “White” means that the gene does not.

Figure 4. Representative variation of gene expression level in lung and Peripheral Blood Mononuclear Cells (PBMCs) for the main pathways involved in the response to OA-lung inflammation across our 24 hours time-course experiment.

We compared the timing and the amplitude of the functional pathways activated in lung and in Peripheral Blood Mononuclear Cell (PBMC) 1h to 24h after Oleic Acid (OA) injection. The X axis represents time-points of the experiment. Numerical values indicate the mean expression level of genes involved in a pathway at a given

time-point.

Figure 5. Experimental design of the experiment.

We studied 37 mice C57BL6/J, female, 7 weeks old. We divided these mice into 6 groups based on the incubation time (1h, 1h30, 3h, 4h, 18h, 24h). In each group, mice are divided into 2 subgroups depending on the injection: Oleic Acid (OA) or physiological serum. 25 samples have been hybridized on 2 microarrays consisting in technical replicates and 12 samples have been hybridized once. We performed 49 hybridizations.

10- Supplemental data

S1: RNA_Agilent_electropherograms.pdf - The quality of total RNA from PBMC isolated using the Mouse RiboPure-Blood RNA isolation kit (Ambion, #AM1951) was confirmed on the Agilent 2100 Bioanalyzer (Agilent, #G2938C) using the RNA 6000 Nano chips (Agilent, #5067-1511). After 1H (A), 1H30 (B), 3H (C), 4H (D), 18H (E) and 24H (F) OA incubation, the electropherograms show that the RNA is neither degraded, nor contaminated with genomic DNA. The X-axis represents the time in seconds and the Y-axis represents the fluorescence intensity (*i.e.* the RNA amount). Peaks at 41.5s and 47s represent 18S and 28S ribosomal RNA respectively.

S2: 1534_differentially_expressed_genes.txt - A set of 1534 differentially expressed genes between across time-course has been identified by SAM with FDR 5%. Each gene is associated with its expression values in OA samples. Each expression value has been normalized by quantile and adjusted with its physiological serum value at the same time point.

S3: inflammatory_process_pathways.txt - Genes differentially expressed across time-course and involved in inflammatory process pathways are grouped into 5 sub-categories: Inflammatory and immune response, DNA damage and DNA repair, Autophagy, Cell adhesion molecules (CAM) and Lipid metabolism. Each gene is associated with its mean expression value at each time point after OA incubation. Data have been normalized using quantile and adjusted on physiological serum expression level at each time point.

11- tables

Table 1. Functional enrichment of genes belonging to cluster A.

Annotation Cluster 2	**Enrichment Score: 1.45**		
Database	*Term*	*Count*	*P-Value*
Gene Ontology	Response to stress	37	1.8E-02
Gene Ontology	Response to endogenous stimulus	18	2.4E-02
Gene Ontology	DNA repair	14	3.4E-02
Gene Ontology	Response to DNA damage stimulus	16	3.9E-02
Gene Ontology	DNA metabolic process	27	9.3E-02
Annotation Cluster 5	**Enrichment Score: 1.14**		
Database	*Term*	*Count*	*P-Value*
Gene Ontology	Nitrogen compound metabolic process	21	1.7E-02
Gene Ontology	Nitrogen compound biosynthetic process	8	2.4E-02
Gene Ontology	Amine metabolic process	19	3.0E-02

Table 2. Functional enrichment of genes belonging to cluster B.

Annotation Cluster 1	**Enrichment Score: 1.9**		
Database	*Term*	*Count*	*P-Value*
Gene Ontology	Regulation of cellular metabolic process	53	7.5E-04
Gene Ontology	Biopolymer metabolic process	89	2.2E-03
Gene Ontology	Transcription	46	5.7E-03
Gene Ontology	Regulation of transcription	44	6.6E-03
Gene Ontology	Transcription DNA-dependent	42	7.0E-03
Gene Ontology	RNA metabolic process	51	7.1E-03
Annotation Cluster 4	**Enrichment Score: 1.09**		
Database	*Term*	*Count*	*P-Value*
Biocarta	IL17 signaling pathway	4	1.7E-02
Gene Ontology	External side of plasma membrane	8	1.8E-02
Gene Ontology	Cell surface	8	9.5E-02
Gene Ontology	Biological adhesion	14	9.8E-02
Gene Ontology	Cell adhesion	14	9.8E-02
KEGG pathway	Cell adhesion molecules (CAMs)	6	9.9E-02
Gene Ontology	Cell-cell adhesion	6	1.7E-01
KEGG pathway	Hematopoietic cell lineage	4	4.2E-01

Table 3. Functional enrichment of genes belonging to cluster C.

Annotation Cluster 2	**Enrichment score: 1.63**		
Database	*Term*	*Count*	*P-Value*
Gene ontology	Immune response	18	1.5E-02
Gene ontology	Inflammatory response	11	1.7E-02
Gene ontology	Chemokine activity	6	2.0E-02
KEGG pathway	Cytokine-cytokine receptor interaction	12	3.1E-02
Annotation Cluster 3	**Enrichment score: 1.55**		
Database	*Term*	*Count*	*P-Value*
Gene ontology	Autophagy	4	1.6E-02
Gene ontology	Autophagic vacuole formation	3	2.8E-02
Gene ontology	Macroautophagy	3	5.0E-02

Table 4. Functional enrichment of genes belonging to cluster D.

Annotation Cluster 2	**Enrichment score: 1.55**		
Database	*Term*	*Count*	*P-Value*
Gene ontology	Cholesterol metabolic process	10	1.2E-03
Gene ontology	Sterol metabolic process	10	1.6E-03
Gene ontology	Steroid metabolic process	12	1.1E-02
Gene ontology	Cellular lipid metabolic process	27	6.50E-002

Figure1

Click here to download high resolution image

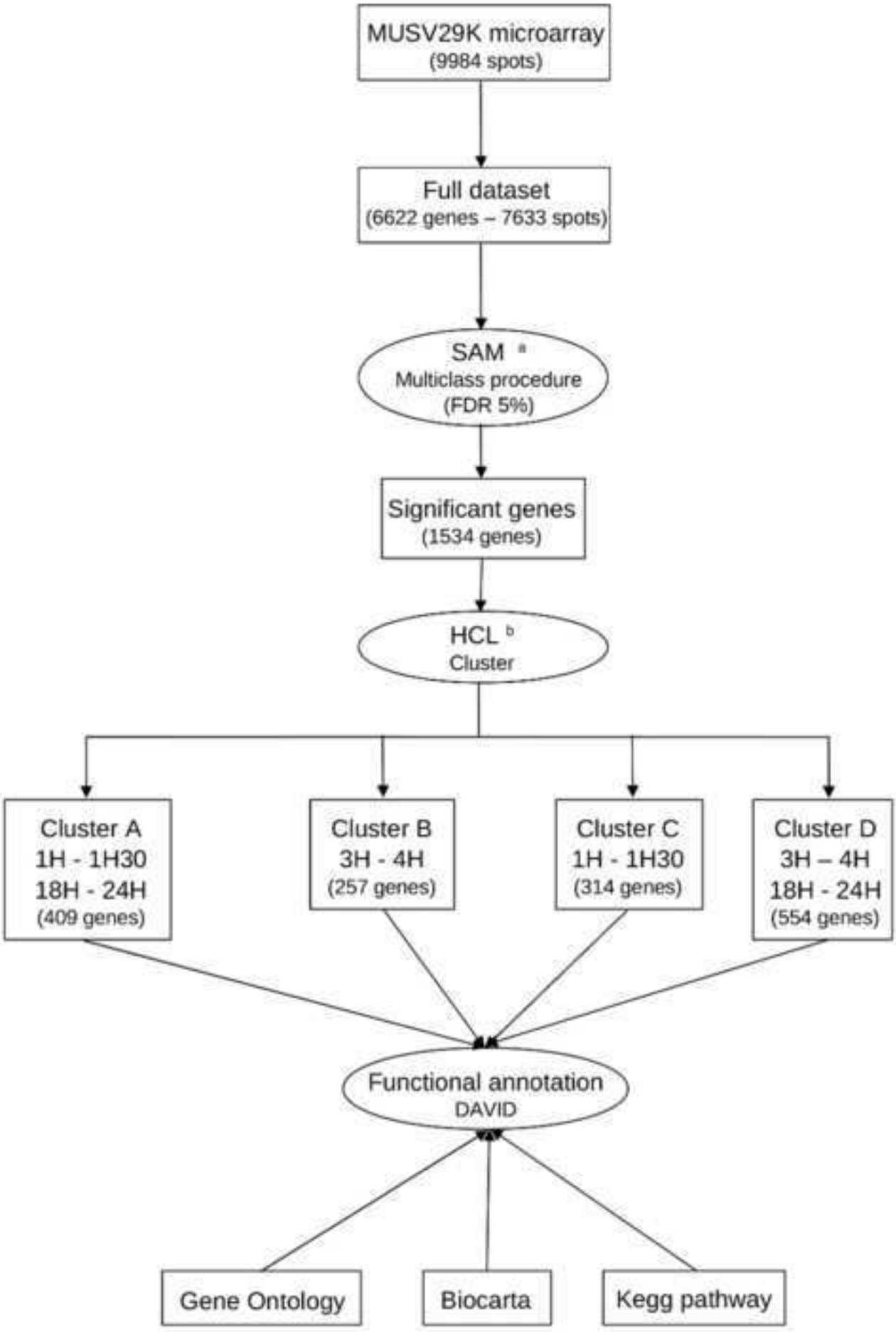


Figure2

Click here to download high resolution image

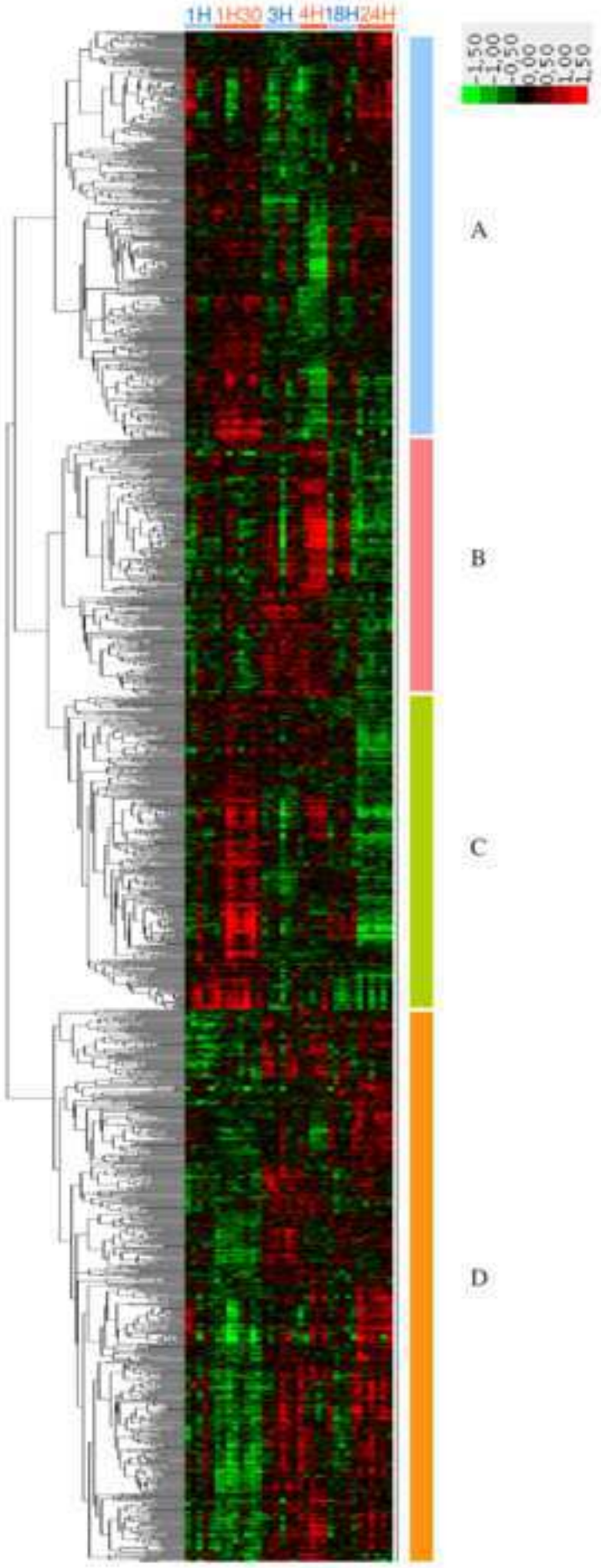


Figure3

Click here to download high resolution image

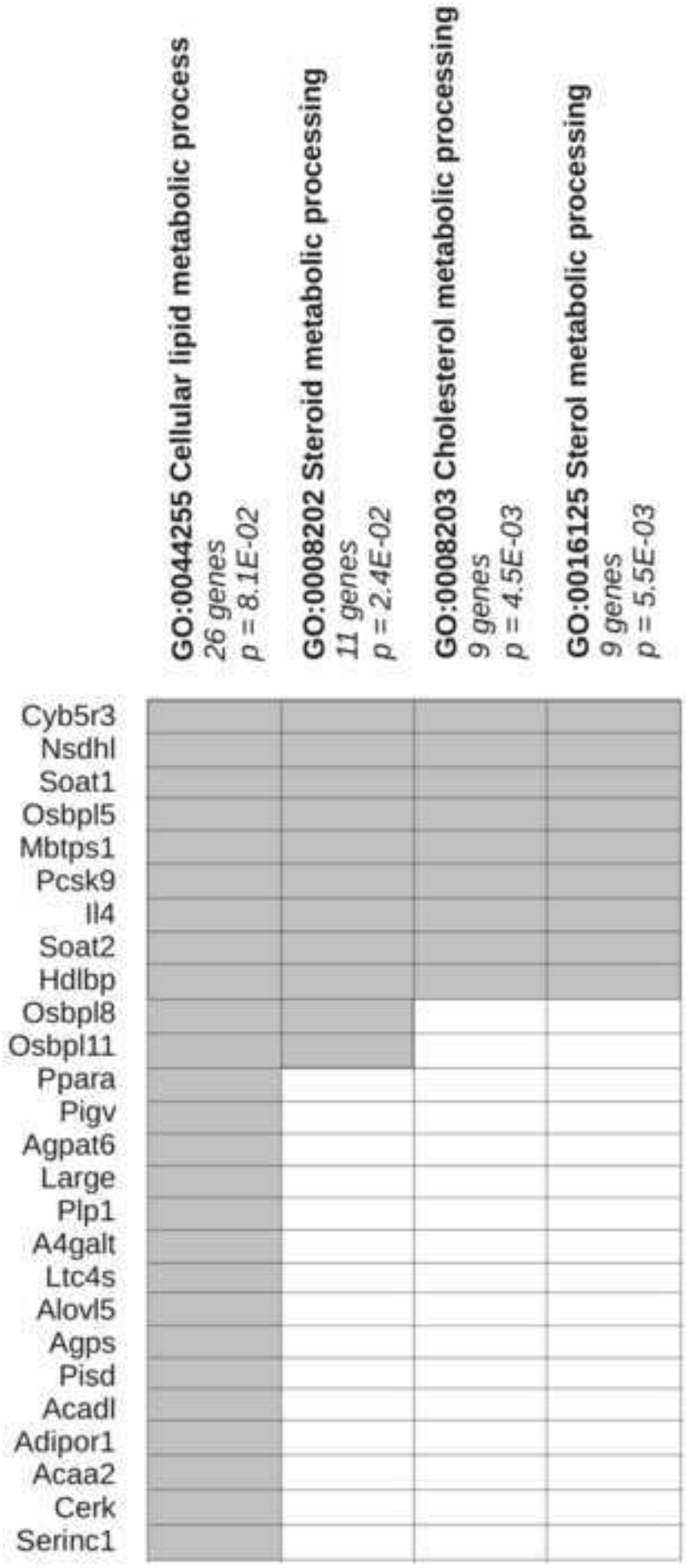


Figure4

Click here to download high resolution image

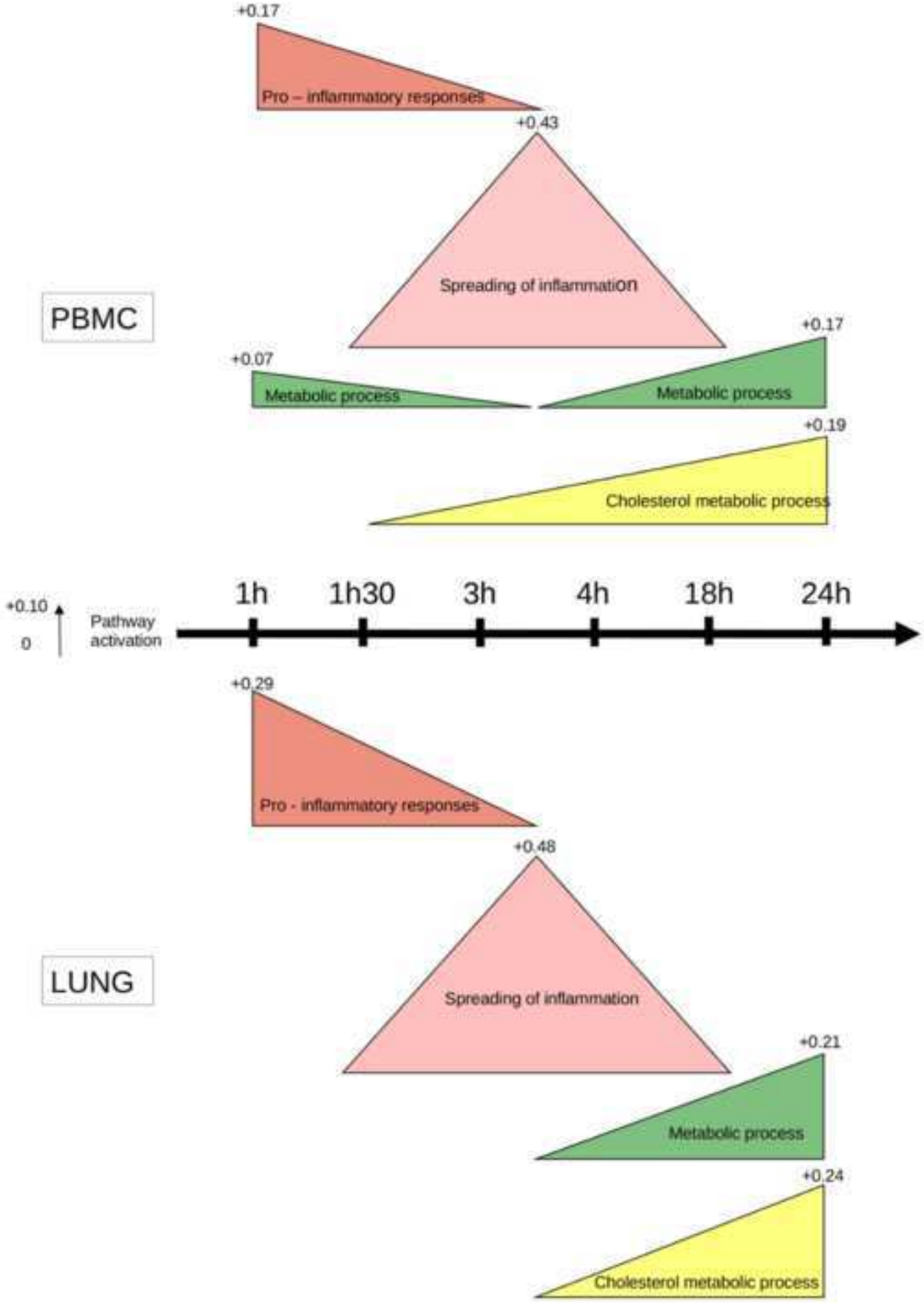
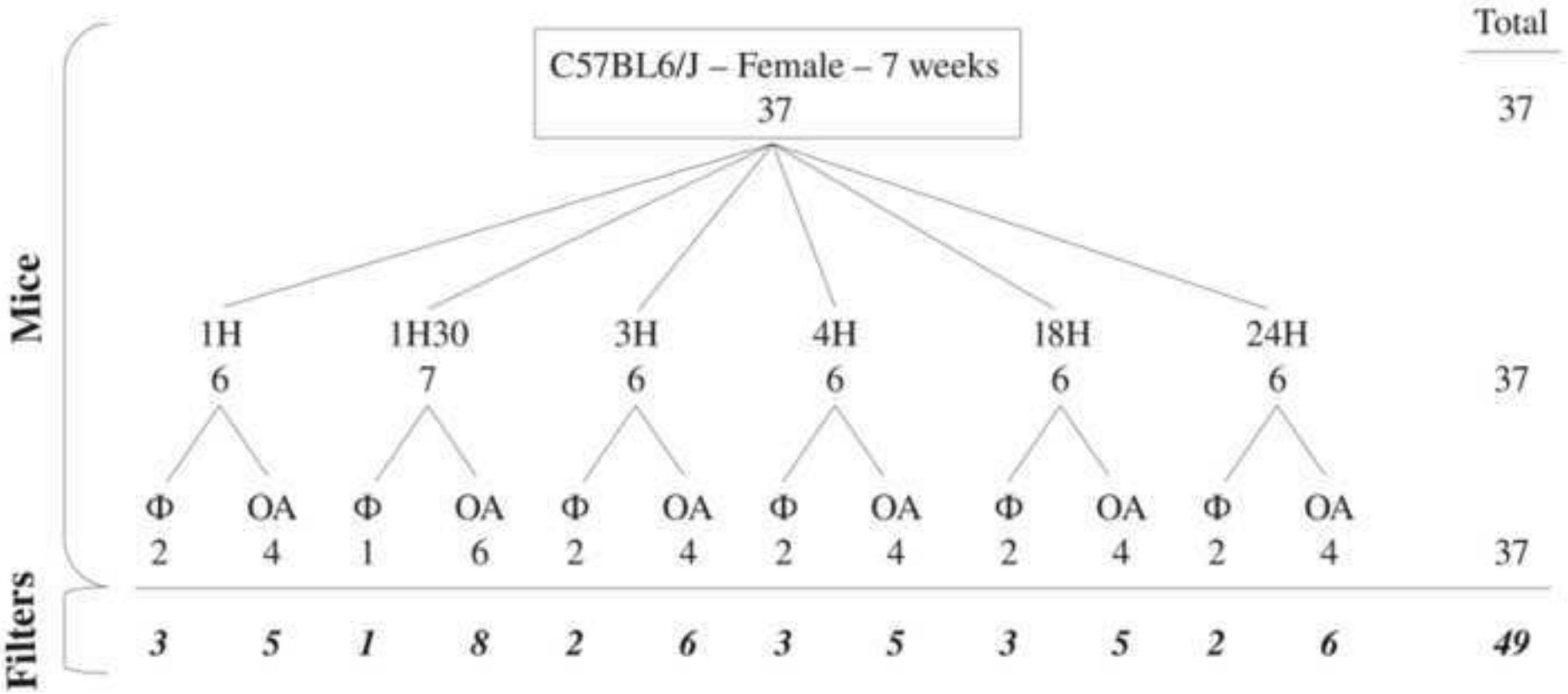
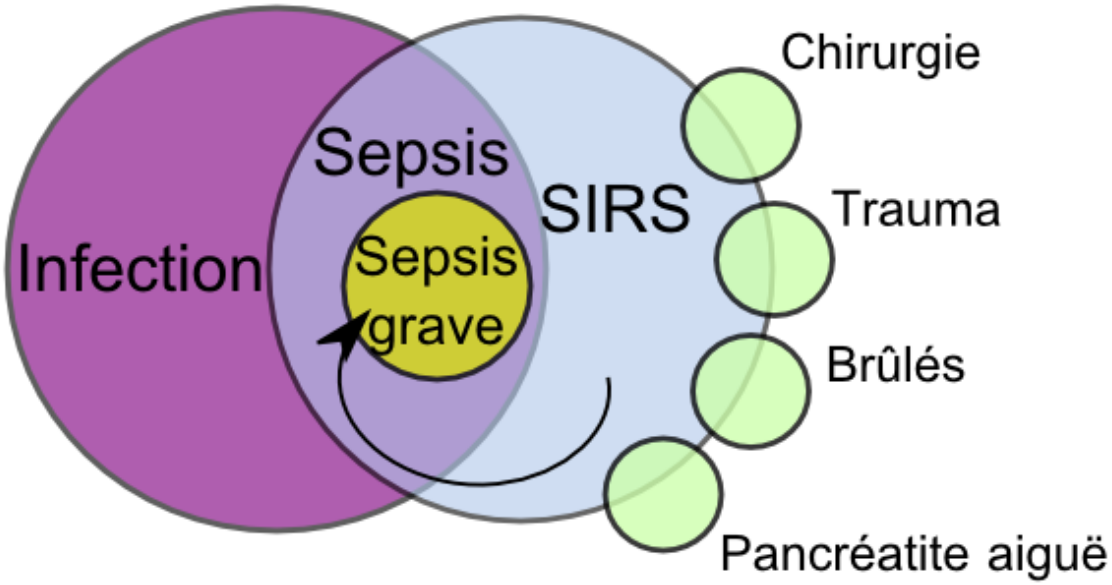


Figure5
Click here to download high resolution image



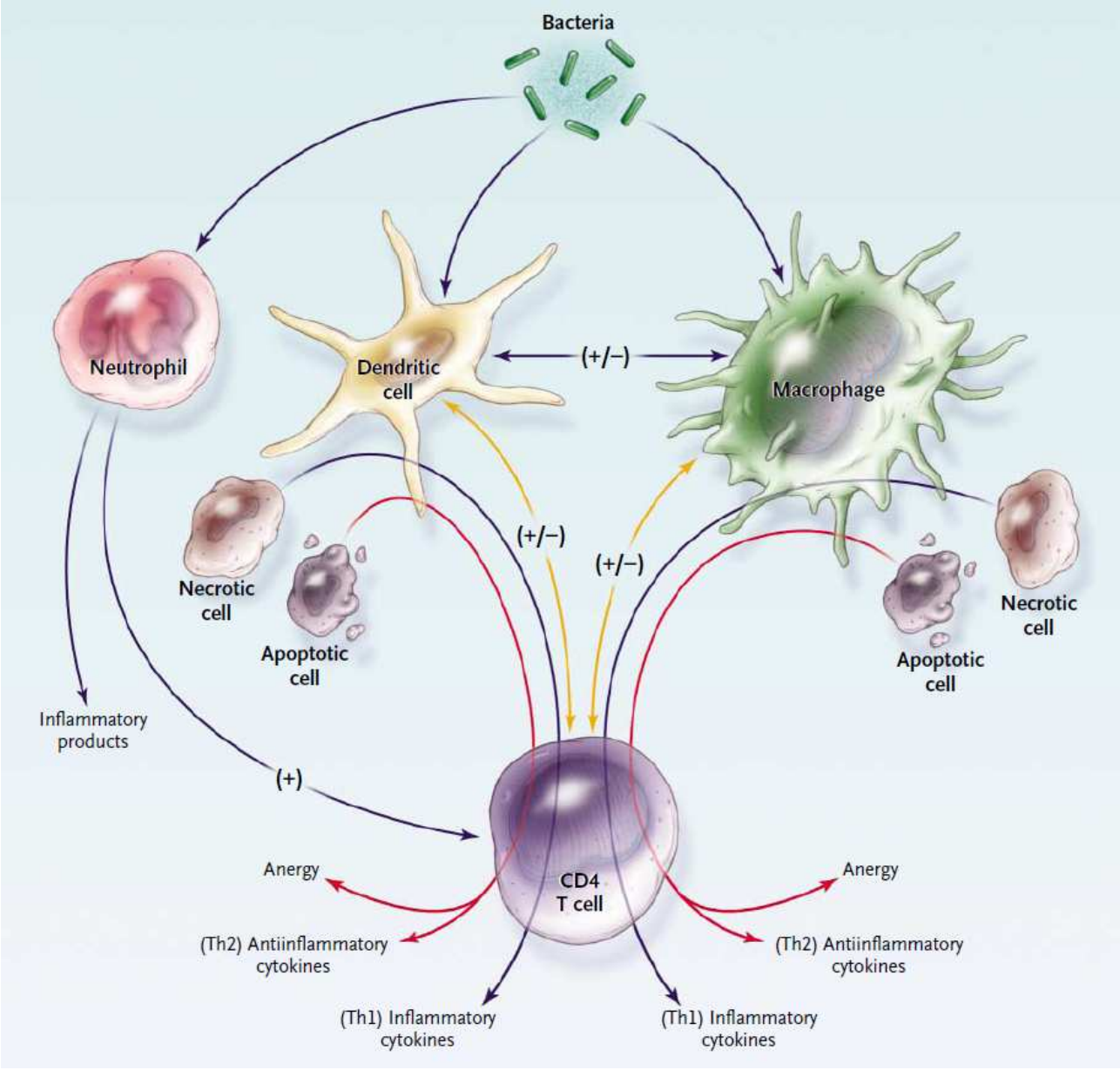
Φ: Physiological serum
OA: Oleic Acid

Figure 1 : Relations schématiques entre les différents états inflammatoires et infectieux



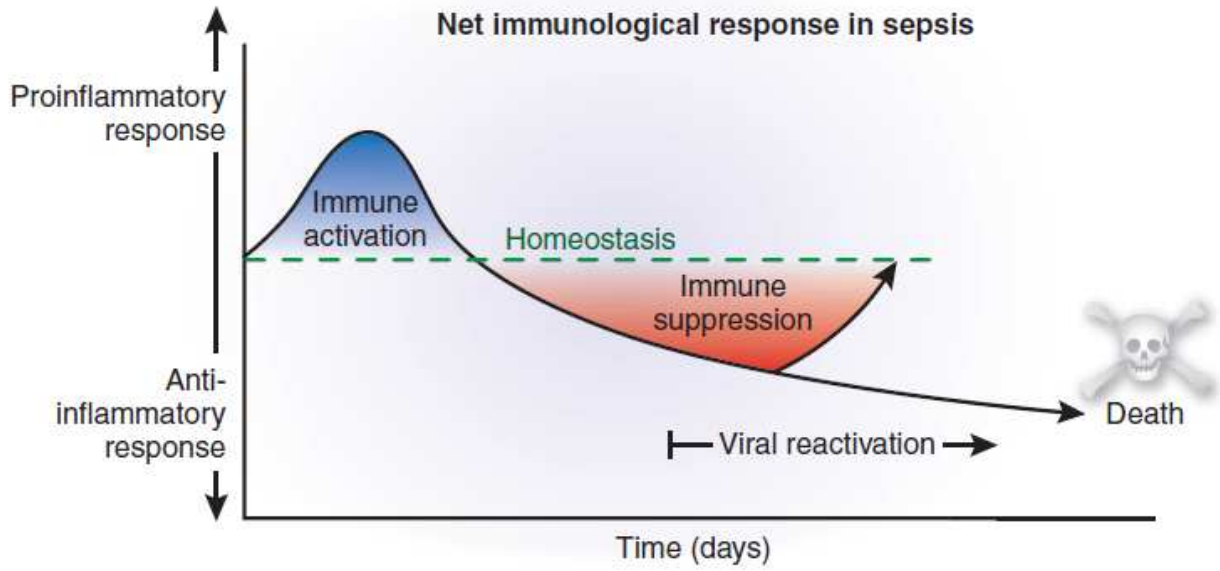
Le syndrome inflammatoire de réponse systémique est le reflet d'une réponse inflammatoire généralisée, dont les étiologies sont variées. Ainsi, la chirurgie, un traumatisme grave, une brûlure étendue ou une pancréatite aiguë induisent une réponse de l'organisme sans qu'il y ait d'infection associée. Le sepsis est une réponse inflammatoire survenant du fait d'une infection. Le sepsis grave associe un sepsis et une défaillance d'organe. Le choc septique est un sepsis grave avec une défaillance hémodynamique nécessitant l'introduction de vasopresseurs.

Figure 2 : Relations entre les cellules du système immunitaire en réponse à l'agression



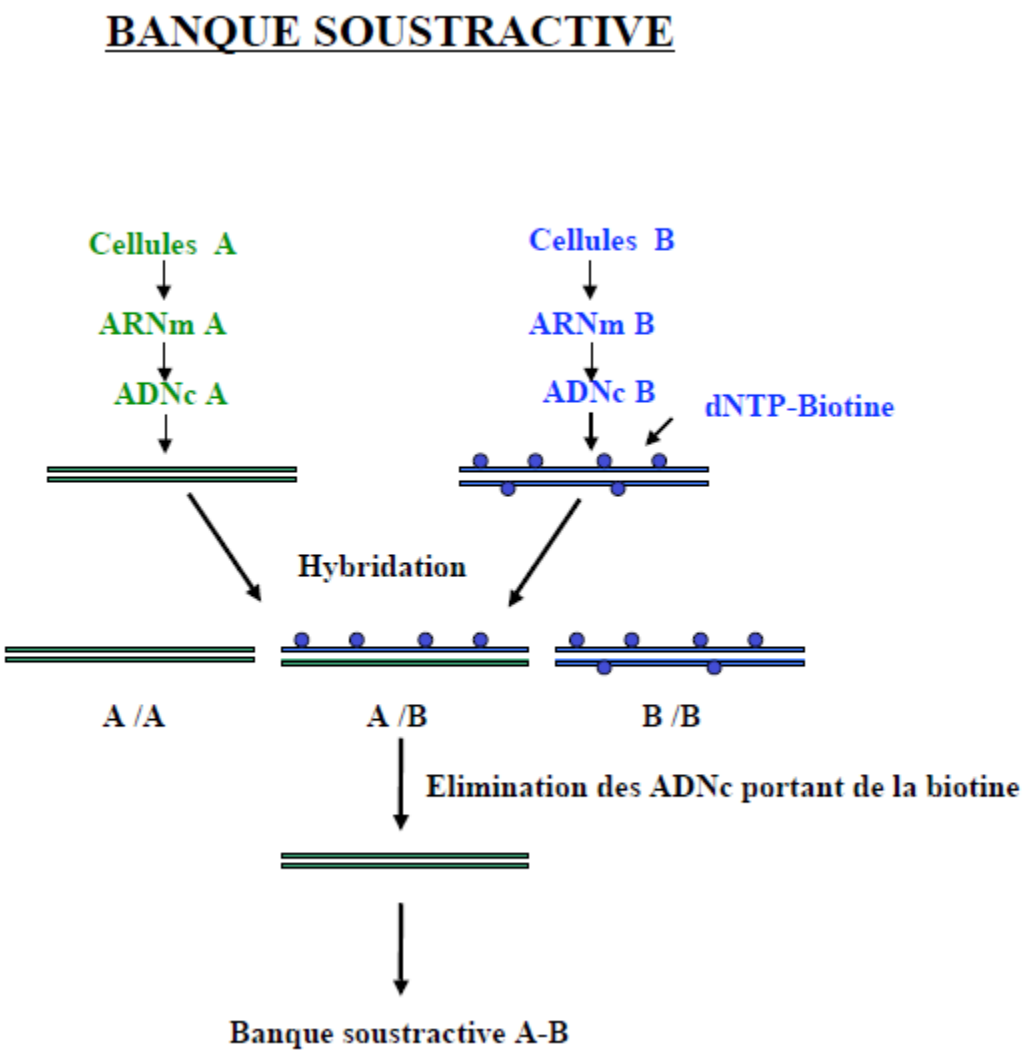
Les macrophages et les cellules dendritiques sont activés par la phagocytose des bactéries et par les médiateurs (IFN-γ par exemple) sécrétés par les lymphocytes T CD4+. Au contraire, les lymphocytes T qui ont un profil anti-inflammatoire (Th2) sécrètent de l'IL-10 qui module l'activation macrophagique. Les lymphocytes T CD4+ sont également activés (type Th1, inflammatoires) via la production d'IL-12 par les macrophages et les cellules dendritiques, et sécrètent en réponse des cytokines pro-inflammatoires. Le dialogue entre les différentes cellules conduit à la modulation de la réponse, soit inflammatoire, soit anti-inflammatoire, voire l'évolution vers un état d'anergie. Les flèches bleues indiquent une activation de type pro-inflammatoire. Les flèches rouges une activation de type anti-inflammatoire. Les signes « + » et « - » symbolisent respectivement « activation » ou « inhibition ». Th1/2 : Lymphocyte T de type helper 1 / 2. Figure reproduite à partir de [3] avec la permission de l'auteur.

Figure 3 : Evolution de la réponse inflammatoire et de ses conséquences lors du choc septique



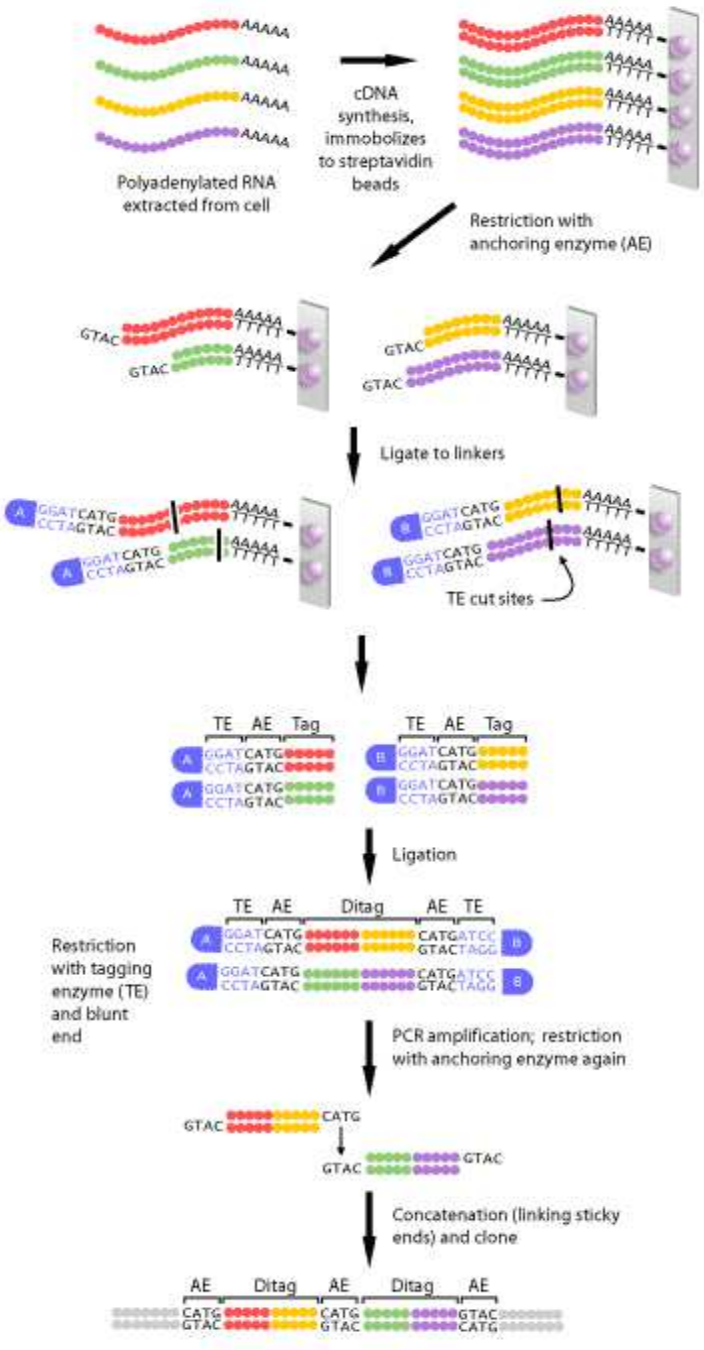
Bien que les deux types de réponses (pro- et anti-inflammatoire) soient activées précocement dans le sepsis, la réponse pro-inflammatoire prédomine initialement. Au cours de l'évolution, la réponse anti-inflammatoire devient prépondérante et c'est durant cette phase tardive que surviennent des infections secondaires (bactériennes ou réactivations virales). Figure reproduite à partir de [5] avec la permission de l'auteur.

Figure 4 : Construction d'une banque soustractive



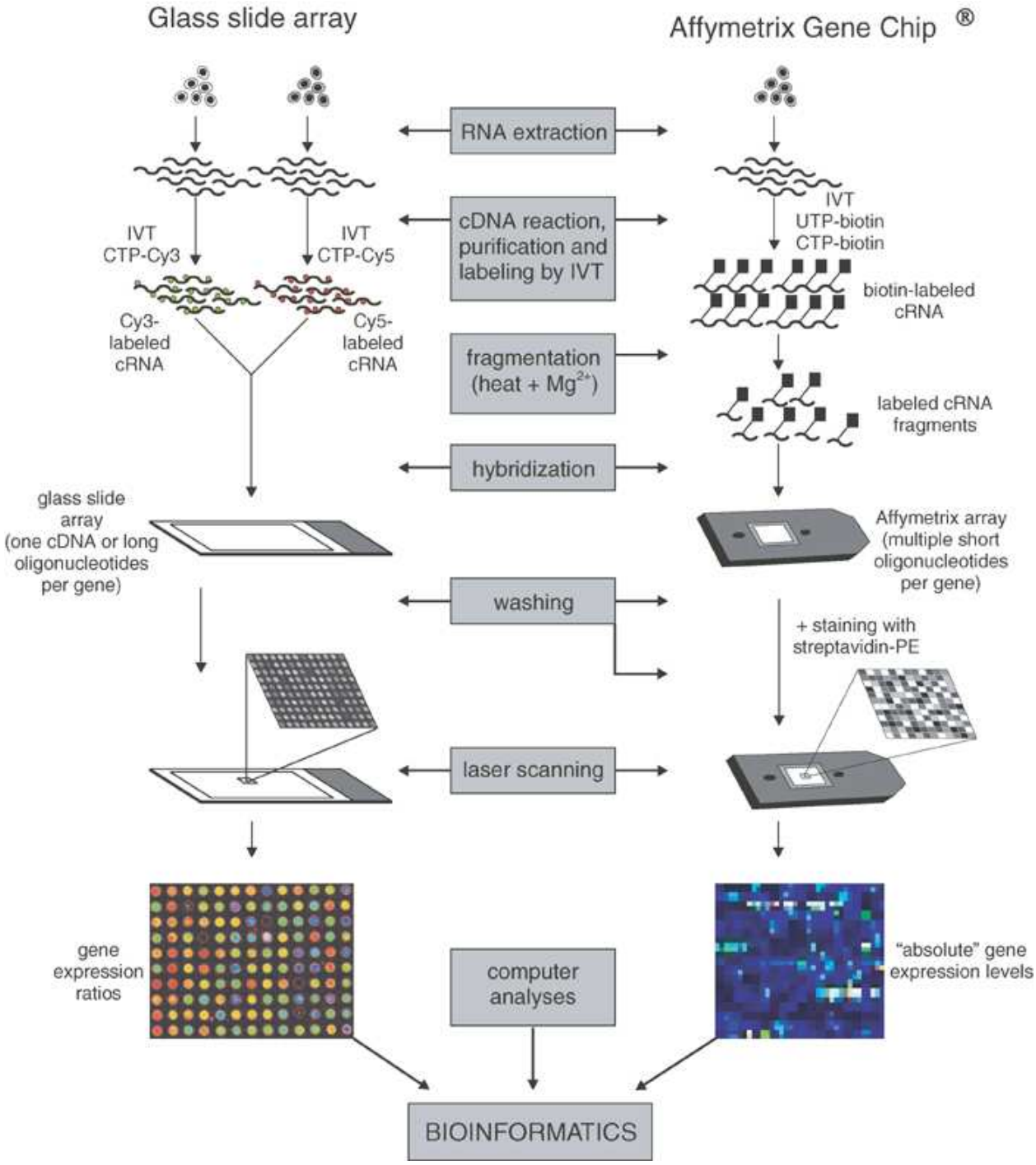
Deux transcriptomes A et B sont extraits de deux conditions expérimentales ou de deux tissus différents. L'un des deux transcriptomes est marqué par de la biotine lors de l'étape de rétro-transcription (ici le transcriptome B). Les ADN complémentaires issus de A et B sont mélangés et mis en condition d'hybridation. Les ADNc marqués sont éliminés (les double-brins d'ADNc de B, ou les ADNc hybrides pour les gènes communs à A et B). Les ADNc spécifiques du transcriptome A sont alors clonés dans des vecteurs d'expression. Figure reproduite avec la permission de l'auteur (Christine Dole, Université de Strasbourg).

Figure 5 : Principe de la méthode « Serial Analysis of Gene Expression » (SAGE)



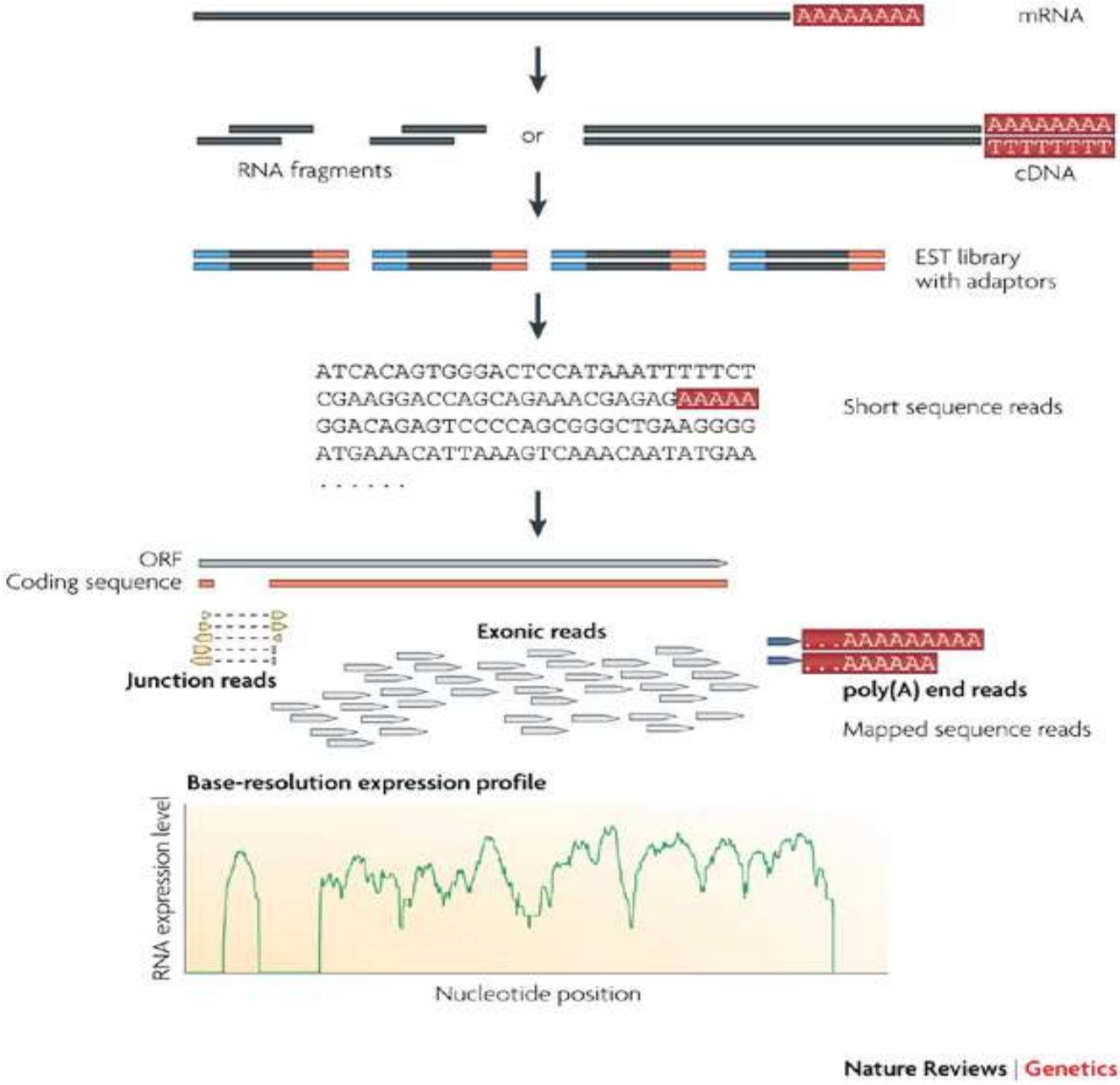
1. Rétrotranscription des ARNm à partir d'une amorce poly-dT couplée à de la biotine. Les ADNc sont purifiés à l'aide de streptavidine, et soumis à une digestion par NlaIII (« anghoring enzyme » (AE)). Tous les fragments possèdent ainsi une extrémité 5'-CATG. 2. Ligation avec un premier adaptateur ("linker") A à l'extrémité CATG. Pour l'autre moitié du prélèvement, ligation avec l'adaptateur B. Ces adaptateurs contiennent des sites de fixation uniques (5'-GGGAC) pour l'enzyme d'étiquetage ("tagging enzyme" TE) BsmFI, et une extrémité compatible avec l'enzyme de restriction NlaIII. 3. Digestion par l'enzyme BsmFI : cette enzyme génère les étiquettes (fragments des ADNc initiaux qui seront séquencés) en coupant 10 nucléotides en 3' de sa séquence de reconnaissance. Les fragments d'ADNc (libérés des billes magnétiques) portent tous la séquence de l'adaptateur à leur extrémité 5'. En revanche, ils diffèrent tous à leur extrémité 3' par les 10 à 14 nucléotides spécifiques de leur étiquette. 4. Ligation des étiquettes par leur extrémité 3' ("ditags"). Les « ditags » sont amplifiées par PCR avec 2 amorces spécifique des adaptateurs A et B. 5. Digestion par NlaIII qui libère les adaptateurs.On isole alors les « ditags » qui sont concaténés (jusqu'à une cinquantaine), clonés et séquencés. Le compte des étiquettes, une fois que celles-ci ont été alignées sur un génome de référence, permet d'avoir une évaluation quantitative de l'expression du gène identifié. Figure reproduite à partir de [18] avec la permission de l'auteur.

Figure 6 : Principe de l'analyse par puces à ADN



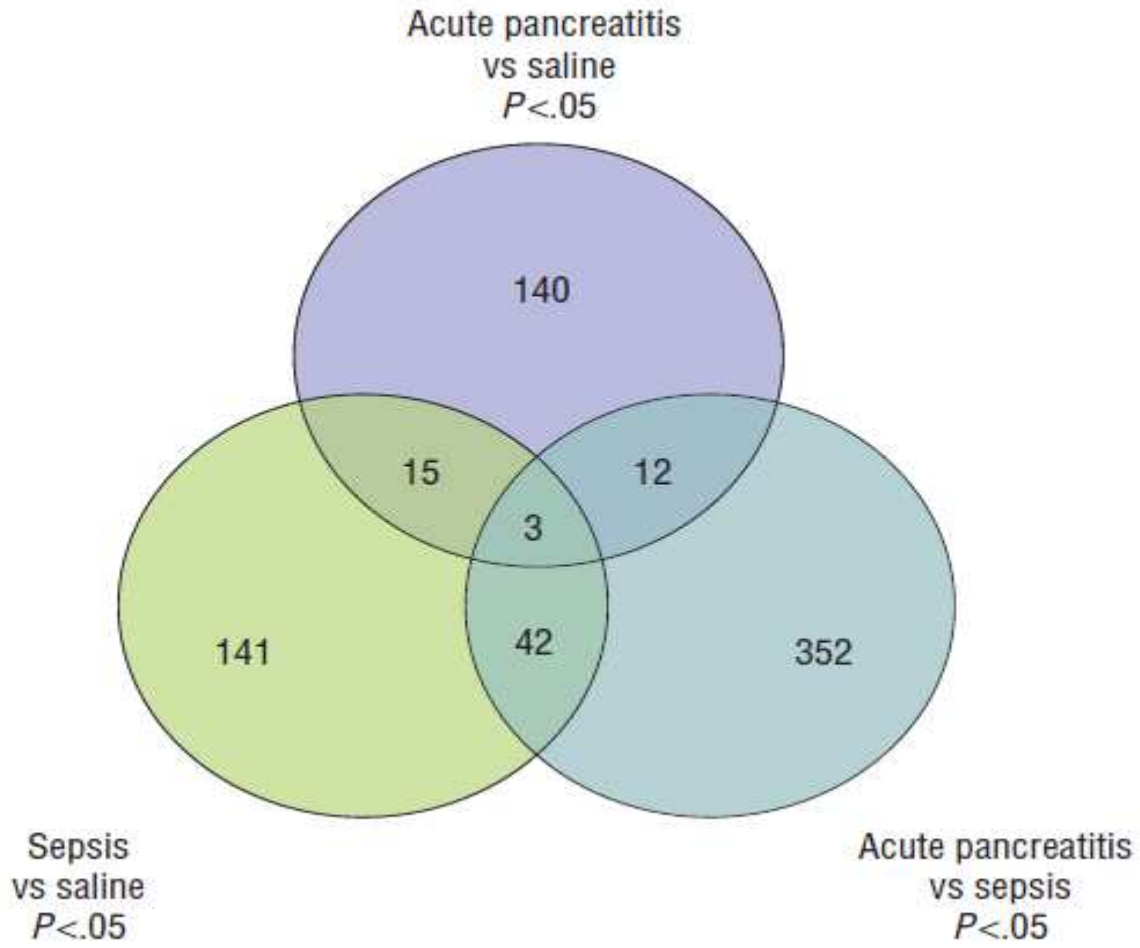
La figure présente deux approches classiques d'analyse du transcriptome par puce à ADN. De manière générale, les ARNm sont rétro-transcrits en ADNc. Au cours de cette étape, les ADNc sont marqués à l'aide d'un fluorochrome ou d'un isotope radioactif. Les ADNc sont hybridés sur les puces, et après un lavage, une image de la puce est générée par un scanner adapté. L'intensité de chaque point de la puce reflète la quantité d'ARNm de chaque gène dans le prélèvement initial. Dans la partie de gauche, on réalise une analyse comparative de deux transcriptomes A et B marqués par deux fluorochromes différents. Cette méthode ne permet d'obtenir que des « ratios d'expression » entre les deux conditions A et B étudiées. Au contraire, dans la partie de droite, un seul transcriptome est hybridé par puce. Cette technique demande deux fois plus de puces à ADN, mais permet de ré-analyser et de comparer des jeux de données différents. Figure reproduite à partir de [19].

Figure 7 : Principe de l'analyse par séquençage haut-débit (RNA-seq)



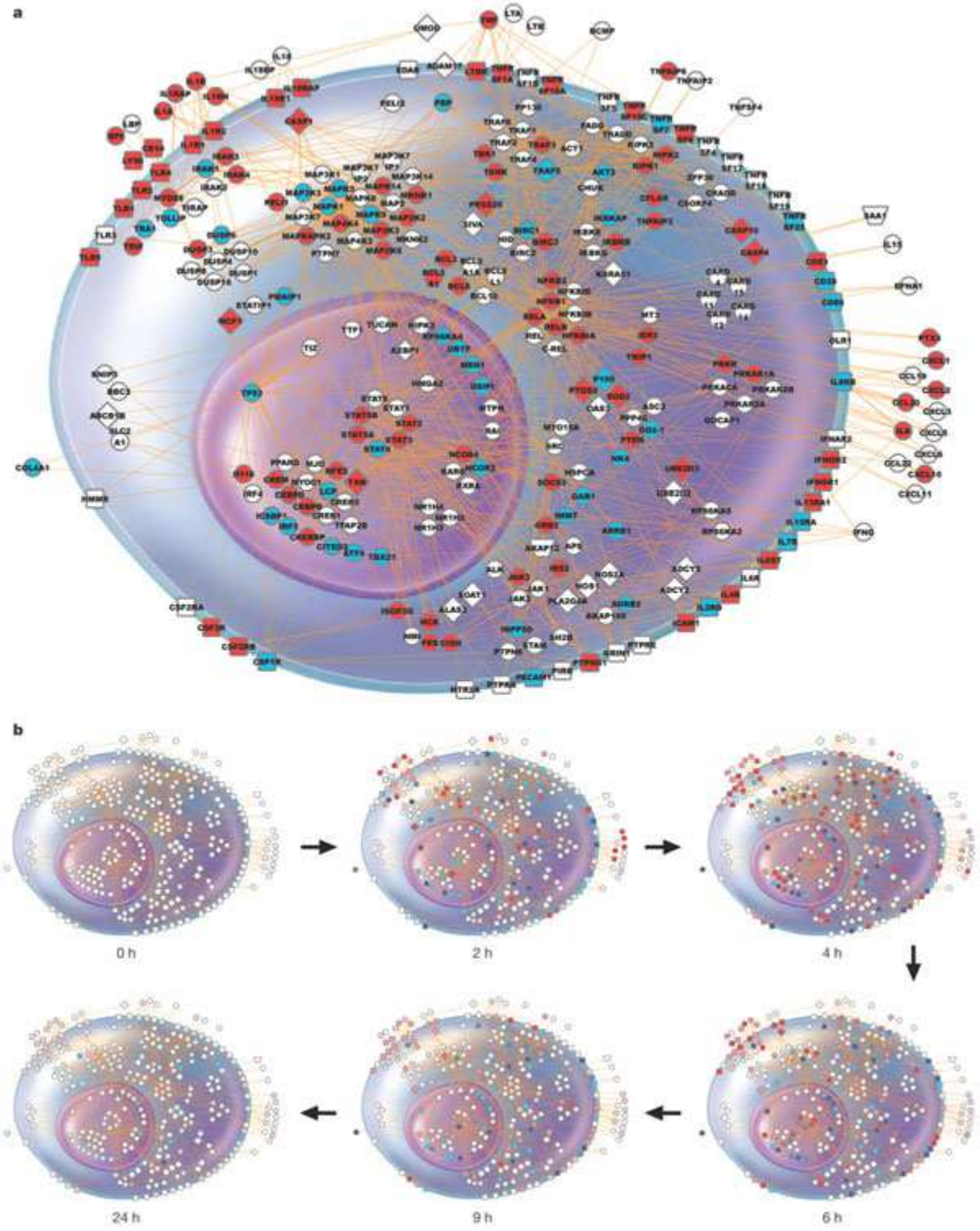
Brièvement, les ARN sont dans un premier temps convertis en une librairie de fragments d'ADNc, soit par fragmentation initiale des ARN, ou fragmentation secondaire des ADNc. Des adaptateurs de séquençage (bleu) sont alors ajoutés aux fragments. La séquence des fragments est alors obtenue par séquençage haut-débit. Les courtes séquences obtenues (« reads ») sont alignées sur le génome ou le transcriptome de référence et classés en séquences exoniques, jonctionnelles ou poly-(A). Comme cela est représenté dans le bas de la figure, ces données d'alignement sont utilisées pour générer un profil d'expression pour chaque gène. Figure reproduite à partir de [29].

Figure 8 : Spécificité des signatures en fonction de l'origine de la réponse inflammatoire



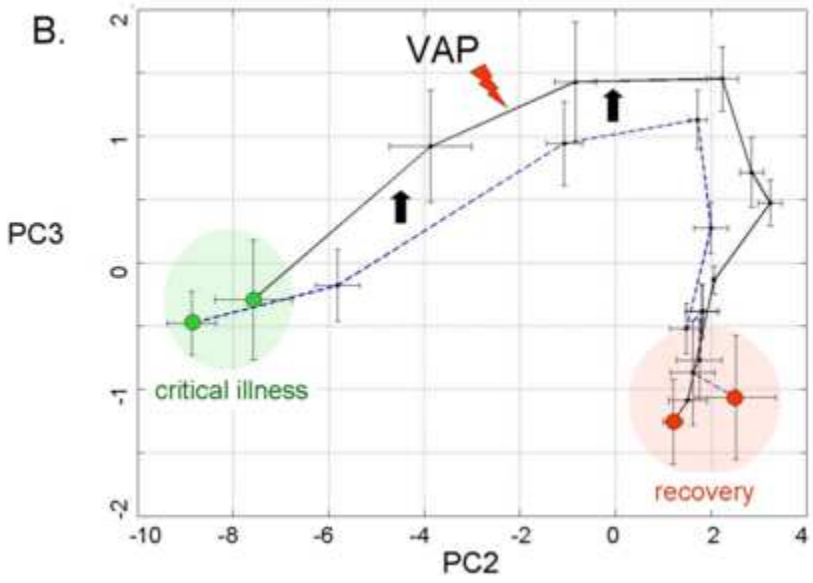
Le diagramme de Venn ci-dessus montre la forte spécificité des signatures transcriptionnelles obtenues lors de l'analyse d'un modèle d'infection et d'inflammation. Les signatures ont été obtenues en comparant deux à deux les modèles inflammatoire (pancréatite), infectieux (infection intra-abdominale) et contrôle. Seuls 15 (5%) gènes sont communs aux signatures spécifiques de l'inflammation d'origine infectieuse ou non. Figure reproduite à partir de [49].

Figure 9 : Evolution du transcriptome après injection de lipopolysaccharide chez des volontaires sains.



Représentation schématique des gènes et voies métaboliques activées (rouge) ou inhibées (bleu) lors de l'injection de lipopolysaccharide chez des volontaires sains. La représentation principale (**a**) est reproduite afin de montrer le rôle fondamental du temps dans la modulation de l'expression des gènes (**b**). Figure reproduite à partir de [67].

Figure 10 : Riboleukogramme



Représentation schématique d'un « Riboleukogramme ». Cette représentation est issue d'une analyse en composante principale de l'expression d'un groupe de 85 gènes au cours du temps. L'expression des gènes était analysée par puce à ADN, à partir de prélèvements sanguins effectués tous les deux jours chez des patients de réanimation soumis à une ventilation mécanique. La construction du riboleukogramme permet de suivre l'évolution des patients (prélèvement initial = rond vert), avec un diagnostic moléculaire précédant le diagnostic clinique (flèche rouge), et une inversion de la variation au moment de la rémission (rond rouge). Figure reproduite à partir de [70] avec la permission de l'auteur.

Tableau 1 : Caractéristiques comparées des différentes techniques de séquençage haut-débit

	454 sequencing	Ilumina	SOLID
Chimie	Pyroséquençage	Polymerase-based	Ligase-based
Méthode d'amplification	Emulsion	Bridge	Emulsion
Taille des fragments	3kb	200bp	3kb
Mb séquencés par analyse	100Mb	1 300Mb	3 000Mb
Durée d'une analyse	7 heures	4 jours	5 jours
Coût par analyse	8 500 $	9 000 $	17 500 $
Coût par Mb	85 $	6 $	6 $